Journal of Experimental Education

A QUARTERLY
Edited by
A. S. BARR

With the Cooperation of

ARTHUR T. JERSILD
PALMER O. JOHNSON
H. H. REMMERS
J. WAYNE WRIGHTSTONE

* * *

BUREAU OF EDNL. & PSYL.
LIBRARY

VOLUME ~~XXIV~~ 25
September, 1956 - ~~June~~, 1957

DEMBAR PUBLICATIONS, INC.
303 East Wilson Street
Madison 3, Wisconsin

TABLE OF CONTENTS

Volume XXV

September, 1956

A Study of the Relationship of Certain Competencies to Success in Teaching Vocational Agriculture: Andrew Pumphrey Torrence.......... 1

The Role of Evaluation in the Improvement of Instruction in Institutions of Higher Learning: James E. Greene.......... 33

The Role of Evaluation in the Improvement of Placement and Follow-Up in Institutions of Higher Learning: Paul I. Clifford.......... 37

The Role of Evaluation in Special Scholarship Selection Programs: Wesley W. Walton.......... 41

The Role of Evaluation in the Development and Improvement of Regional Programs of Higher Education: A. J. Brumbaugh.......... 43

The Role of Sixteen Millimeter Motion Pictures and Projected Still Pictures in Science Unit Vocabulary Learning at Grades Five, Six, and Seven: Louis Romano.......... 49

Relation Between Enrollment of High Schools from Which Students Graduated and Academic Achievement of Agricultural Students, A. and M. College of Texas: John R. Bertrand.......... 59

Children's Thinking in the Process of Subtraction: E. Glenadine Gibb.......... 71

An Experimental Study of Case Discussion Learning: Leonard A. Ostlund.......... 81

Do Probationary College Freshmen Benefit from Compulsory Study Skills and Reading Training? Walter S. Blake.......... 91

December, 1956

A Study of the Relationships of Certain Environmental Factors to Teaching Success: William B. Knox.......... 95

Analysis of Certain Aspects of Teaching Ability: Margaret Lois Jones.......... 153

March, 1957

Computational Illustrations of the Internal and External Consistency Analysis of Examination Responses: William J. Moonan 181

The Relative Achievement of the Objectives of Elementary School Science in Minnesota Schools: Jane Johnston 191

Controlled Experimentation in the Classroom: Julian C. Stanley 195

Recent Techniques for Analyzing Association in Contingency Tables as Applied to Analytical Follow-Up Survey of Education Graduates: Samuel T. Mayo 203

The Best Linear Estimate of the Predicted Value and the Standard Error of the Estimate: Palmer O. Johnson 233

The "Equating" of Non-Parallel Tests: William H. Angoff 241

The Comparability of the Bi-Factor and Second-Order Factor Patterns: John Schmid 249

An Analysis of Variance of Multiple Measurements on Subjects Classified in Unequal Groups of One Dimension: Raymond O. Collier and Clayton L. Stunkard 255

June, 1957

A Program for the Selection of Trainees in Physical Medicine: Nick John Topetzes 263

Testing the Assumptions Underlying Multiple Regression: Francis A. C. Sevier 323

The Interrelationship of Vocabulary, Comprehension and Rate Among Disabled Readers: James C. Reed and Roger S. Pepper 331

A Quick and Dirty Method for Estimating Multiple Correlations: William J. Moonan 339

Journal of Experimental Education

A STUDY OF THE RELATIONSHIP OF CERTAIN COMPETENCIES TO SUCCESS IN TEACHING VOCATIONAL AGRICULTURE

ANDREW PUMPHREY TORRENCE
Tuskegee Institute

SECTION I

Statement of the Problem

THE PURPOSE of this study is to ascertain the relationship between the success attained by vocational agriculture teachers and certain competencies that these vocational agriculture teachers possess. In order to facilitate the achievement of this purpose, four null hypotheses have been formulated. These hypotheses serve as a guide in planning the investigation, in determining what data to collect, in classifying and ordering the data, and in making the basic judgments necessary in this study. The four null hypotheses to be tested in this study are:

1. There is no statistically significant relationship, as shown by coefficients of correlation, between a vocational agriculture teacher's knowledge of technical agricultural information and his success as a teacher.
2. There is no statistically significant relationship, as shown by coefficients of correlation, between a vocational agriculture teacher's proficiency in performing agricultural manipulative skills and his success as a teacher.
3. There is no statistically significant relationship, as shown by coefficients of correlation, between a vocational agriculture teacher's knowledge of professional educational information and his success as a teacher.
4. There is no statistically significant relationship, as shown by regression equations, between a vocational agriculture teacher's knowledge of technical agricultural information, proficiency in performing agricultural manipulative skills, and knowledge of professional educational information and his success as a teacher.

Definition of Terms

The definition of words and phrases given below will generally apply wherever these words and phrases appear throughout this study. Wherever any of these terms are used in this study with meaning different from the definitions here given, the reader will be so informed.

1. Competence—the level of efficiency, ability, or knowledge that one has achieved.
2. Manipulative skills—the ability to use one's knowledge and intellect to perform tasks with the hands or by mechanical means.
3. Null hypothesis—a supposition that observed effects are due to chance factors and not to statistically significant relationships.
4. Professional education—that part of a teacher's training that will give him an appreciation of and a knowledge of how best to teach his subject and influence desirable changes in his students. Professional education would concern itself with such things as the place of the school in our society, child growth and development, child psychology, guidance and measurement in education, and principles and methods of teaching.
5. Relationship—a connection or link between two or more factors. It means that alterations in one or more factors occur along with alterations in one or more different factors.

Ednl. Psy. Research
DAVID H. ... ING COLLEGE

---

*The author wishes to express his sincere thanks and appreciation to many people whose assistance has been sought and secured in the development of this study, particularly Dr. W. T. Bjoraker, his major professor, Dr. B. W. Kreitlow, and Dr. A. S. Barr for constant direction, encouragement, and inspiration throughout the course of this investigation.

These changes may occur conversely, or they may occur in harmony with each other.

6. Teaching success—the degree to which a teacher of vocational agriculture uses desirable means to attain desirable ends in discharging his duties as a teacher.
7. Teacher training—those activities needed for assisting teachers or prospective teachers to secure the professional knowledge, abilities, understandings, and appreciations which will enable them to qualify for professional employment or advancement in teaching vocational agriculture. This training is generally given in accredited colleges or universities where agricultural education is taught.
8. Technical agriculture—the knowledge that one has concerning factual information on the production and management of livestock, crops, and soil.
9. Vocational agriculture—a nation-wide, federally aided program of systematic instruction in agriculture and farm mechanics of less than college grade conducted in public schools or classes for persons over 14 years of age, for the purpose of training present and prospective farmers for proficiency in farming. This program in the individual states and schools must be conducted in conformity with Congressional Vocational Education Acts. (1)

Background of the Study

The program of vocational education of less than college grade in public schools of the United States is of relatively recent development. It was developed in conformity with the provisions of Public Law No. 347 - Sixty-fourth Congress, approved February 23, 1917, usually referred to as the Smith-Hughes Act. (1) Supplementary acts have been enacted from time to time. The latest of these is the Vocational Education Act of 1946, commonly known as the George-Barden Act. These acts have been enacted for the purpose of promoting and developing vocational education through a plan for cooperation between the federal government and the states. Vocational education in agriculture was encompassed in the first enabling act and it has been recognized and included in each succeeding act.

This plan of cooperation for the development of vocational education is based upon two fundamental ideas: (1) that vocational education is a matter of national interest and essential to the national welfare, and (2) that federal funds are necessary to stimulate and assist the states in making adequate provisions for such training. (1)

The primary aim of vocational education in agriculture is to train present and prospective farmers for proficiency in farming. It has been considered as one of the permanent phases of the public school program in rural areas. Vocational education in agriculture has been instrumental in developing the skills, abilities, information, attitudes, understandings, and working habits of rural people which will enable them to enter employment in agriculture, or to make satisfactory advancement in farming occupations. It also, presumably, makes a contribution toward the development of good citizens, including their health, social, civic, and cultural interests.

Under a plan of cooperation between state boards for vocational education and the United States Office of Education, vocational education in agriculture is a nation-wide, federally aided program of systematic instruction in agriculture and farm mechanics of less than college grade conducted in public schools or classes for those persons over 14 years of age who have entered upon or who are preparing to enter upon the work of the farm or of the farm home. The instruction is given by qualified teachers of vocational agriculture who are employed on a 12-month basis so as to follow-up their instruction throughout the year by assisting individuals and groups with their farming programs on the individual student's farm. This provides a year-round, uninterrupted program of instruction in agriculture.

The major objectives of vocational education in agriculture are to develop effective ability to:

1. Make a beginning and advance in farming.
2. Produce farm commodities efficiently.
3. Market farm products advantageously.
4. Conserve soil and other natural resources.
5. Manage a farm business.
6. Maintain a favorable environment. (1)

There are three well-recognized groups to be served by vocational agriculture, namely,

1. In-school farm youth preparing to farm.
2. Out-of-school young men engaged in farming, usually on the farm or as a farm worker looking forward to full establishment as an independent farm operator.
3. Adult farmers fully established as operators either as owners or as tenants.

Vocational education in agriculture has been considered as an integral part of public secondary school education and contributes to the general objectives and philosophy of such education. An examination of any recognized set of secondary school objectives with the intention of ascertaining what contributions vocational agriculture makes to them will readily prove the authenticity of this claim. Cook (13) postulates how vocational education in agriculture provides many opportunities for the development of abilities essential to the seven cardinal principles of education. This writer made a study of "How the Vocational Agriculture Curriculum of Wisconsin Contributes to the Ten Imperative Needs of Youth as Listed by the Education Policies Commission." (45) He charted the needs of youth which various vocational agriculture subjects help to satisfy, and found that the vocational agricultural curriculum of Wisconsin makes a very appreciable contribution to satisfying these ten imperative needs. All of the imperative needs get some contribution from vocational agriculture subjects, varying from a very little to considerable. The methodology of teaching, the knowledge of the subject, and the teacher himself are major factors in getting the imperative needs of youth satisfied.

Vocational agriculture has served to elevate the level of economic stability, social stability, and general establishment of farm people in farming by teaching agriculture to high school farm boys, as well as by keeping the adult farmers aware of the latest and best scientific discoveries in farming. Without the instructional aid of this program, many young men would not have been able to begin or continue in farming successfully. The more successful vocational agriculture teachers are, the more successful farmers and farm communities will be in attaining desirable ends. This study is intended to reveal what relationship certain competencies have to the success of vocational agriculture teachers.

One of the essential features is a good program of vocational education in agriculture is an adequate supply of competent teachers. This was recognized in the Smith-Hughes Act which provided specific funds for the training of teachers and required that in order to use the money provided for the salaries of vocational teachers, a state must expend a specified amount for the training of teachers. The George-Barden Act does not make a separate authorization of funds for this purpose, but it does state that the funds made available for the several vocational fields (which includes vocational agriculture) may be used for assisting the States and Territories in maintaining adequate programs of teacher training. (1)

Why some vocational agriculture teachers meet with great success while others attain only mediocre success or fail is of great importance to teachers themselves, to state supervisors of vocational agriculture, to teacher training institutions, to high school vocational agriculture students, and to farmers. If the success of vocational agriculture teachers could be successfully predicted before or shortly after he enters the field of teaching, it would save the time, energy, and resources of the teacher and all who help train him. More important, it would protect farming areas from ineffective vocational agriculture teachers.

The number of recent studies attacking various phases of vocational agriculture teacher's success gives evidence of the gravity of this problem.* It is easily recognized that unless vocational agriculture teachers are effective, the whole program of vocational agriculture can be of little value. Some studies concerned with causes for discontinued departments of vocational agriculture in high schools were reviewed by this writer. Three of these studies (35, 36) listed ineffective teaching as being one of the main causes for the vocational agriculture program being dropped from the high school curriculum.

In order to increase the effectiveness of vocational agriculture teachers, we must first discover on what their success hinges. Many of the studies reviewed by this writer deal with pre-service training of teachers of vocational agriculture.. Although this study will reflect the significance of pre-service training, it is not primarily a pre-service training study. It deals directly with the in-service teachers and therefore is considered a contribution that has not been made previously.

In an article entitled "Impact of Research in Agriculture," Byron T. Shaw (34) points out that agriculture faces a tremendous job in the next 25 years. He states that "our population is now growing at the rate of $2\frac{1}{2}$ million a year, and the Census Bureau estimates that by 1975, under moderately favorable conditions, it will reach 190 million. This means that American farmers will have to produce about 30 percent more by 1975 if they are to provide adequately for all our people." Most good land is already taken, so our greatest opportunity for meeting this increased population problem is in better use of the land we already have for farming purposes. This assumes greater efficiency in production on farms and improved distribution, utilization, and marketing after products leave the farm. Mr. Shaw writes, "The job ahead of us is to increase production wherever possible, to prevent declines in production, to cut waste and spoilage of goods during handling, and marketing, and in the home, and to make better use of agricultural products".

---

*See Chapter II, "Review of Related Research," in original thesis on file in the Library of the University of Wisconsin.

This article is principally concerned with research in technical agriculture and technical farming, however, we can easily infer from it the need for efficient teachers of vocational agriculture. Vocational agriculture teachers, along with other rural education agencies, have the tremendous responsibility and obligation not only to funnel the results of agricultural research to every farmer and farm home, but also to interpret these results so that they might be understood, appreciated, and profitably utilized.

In 1900 each American farmer produced enough to feed eight people. Today he produces sufficiently to feed 16 people, but it is estimated that by 1975 he will have to produce enough to feed 21 people. The successful vocational agriculture teacher will do much to help the farmer accomplish this task. It is hoped that as a result of this study, we will have some indications of whether or not certain competencies of vocational agriculture teachers are related to their success.

Aspects of the Problem to be Investigated

While the purpose of this study is to determine the relationship between the success attained by vocational agriculture teachers and certain competencies possessed by these teachers, more specifically, it will seek to find answers to the following questions:

1. What relationship exists between a vocational agriculture teacher's knowledge of technical agricultural information and his success as a teacher?

2. What relationship exists between a vocational agriculture teacher's proficiency in performing agricultural manipulative skills and his success as a teacher?

3. What relationship exists between a vocational agriculture teacher's knowledge of professional educational information and his success as a teacher?

4. What relationship exists between a vocational agriculture teacher's knowledge of technical agricultural information, proficiency in performing agricultural manipulative skills, and knowledge of professional educational information as a composite and his success as a teacher?

Significance of the Study

A perusal of the literature indicates that the competencies of vocational agriculture teachers that are being investigated in this study have not been assessed previously in relation to the success of vocational agriculture teachers. Little is known as to why some vocational agriculture teachers seem to be more successful than others. It is hoped that this study will reveal what relationships the above competencies have on the success of these teachers. This study should be a valuable tool for determining pre-service and in-service training needs of vocational agriculture teachers. It will show whether or not there are wide-spread incompetencies of certain technical information, manipulative skill abilities, and professional knowledge of vocational agriculture teachers in Wisconsin, and how they relate to the success of these teachers.

Wisconsin has long been regarded as a state of high achievement in agriculture and agricultural education. It is hoped that this study will set a pattern of procedure that can be followed in other states, and that it will permit other states making such a study in the future to compare their results with the findings of this study.

Lastly, this whole field of vocational agriculture teacher success is still in need of much serious research. This study will serve as a forerunner for subsequent vocational agriculture teacher-success studies to be conducted at the University of Wisconsin.

SECTION II

DESIGN OF THE STUDY

THE COMPETENCIES of vocational agriculture teachers that are to be measured in this study are in the areas of technical agriculture, manipulative farm skills, and professional education.

The group to be studied will be a random sample of sixty vocational agriculture teachers who teach in one man vocational agriculture departments of public high schools located in approximately the southern two-thirds of the state of Wisconsin.

Teacher success will be determined from ratings given to the vocational agriculture teachers by people who are in a position to know of the work of these vocational agriculture teachers. These ratings will be furnished by state supervisors, principals, pupils, and farmers in the local school and community in which each teacher teaches.

Description of the Subjects

The subjects used in this investigation were a random sample of sixty high school teachers who teach vocational agriculture in one teacher departments located in approximately the southern two-thirds of the state of Wisconsin. More specifically, these teachers were drawn from the area between the upper limits of the county lines of St. Croix, Dunn, Chippewa, Clark, Marathon, Shawano, Brown, and

Kewaunee counties and the southern limits of the state of Wisconsin. (See Figure 1.) The total population from which these teachers were drawn did not include vocational agriculture teachers who teach in vocational schools, and it did not include vocational agriculture teachers who teach in departments of vocational agriculture that have two or more teachers.

All of the vocational agriculture teachers who taught in one-man departments located within the area described above were listed alphabetically by name. Numbers were assigned to this alphabetical listing, and a sample of 60 teachers were drawn by the use of Fisher and Yates' table of random numbers. (19) There were 217 teachers in the total population.

Out of the sixty teachers chosen at random, 16 were currently enrolled in summer school at the University of Wisconsin. There were 14 other vocational agriculture teachers enrolled in summer school at the University who taught in one-man departments located in the area described above but those assigned numbers did not fall in the random selection. It was decided that if these 14 teachers did not differ significantly from the 16 teachers who were drawn in the random sample, they too could be used in the study as a part of the random group.

Pursuant to that decision, the two groups were analyzed with respect to the following: (1) age, (2) training, (3) rearing (whether or not teachers were reared on a farm), (4) experience in teaching agriculture, (5) total student enrollment in high schools in which teachers taught, and (6) total vocational agriculture department enrollment in high school in which teachers taught. The results of this analysis are shown in Table I.

The training of these teachers is very similar. They all have a B. S. degree in agricultural education or its equivalent, and have done some post graduate work. One of the teachers has a M. S. degree in agricultural education. All of the teachers except one were reared on a farm. The one who was not has owned a farm in adult life.

The statistical significance of the difference between means (24) for the two groups was calculated for the age, experience in teaching agriculture, total high school enrollment, and total vocational agriculture department enrollment columns. It was decided that if these differences were not significant at or below the .05 level of probability, it could be assumed that these two groups were a part of the same total population. All differences proved non-significant with the greatest difference being at the 15% level of probability for the total high school enrollment column. Other differences were at the 90% level for age, the 65% level for experience in teaching agriculture, and the 40% level of probability for the total vocational agriculture department enrollment factors.

Since the differences in these two groups are not significant at the levels described above, and since the 16 teachers were randomly drawn from the total population, it was assumed that the 14 teachers who were in summer school but not drawn in the random selection could be considered a part of the randomization, and they were therefore added to the subjects used in this study. This gave a total of 30 teachers on the University campus to be used in the study, and it left 30 additional teachers to be selected who were not enrolled in summer school at the University of Wisconsin. The remaining 30 teachers participating in this study who were not enrolled in summer school were chosen as their assigned numbers appeared in the random selection providing they were available for the testing program designed for this study. All of the 30 teachers in the state whose assigned numbers were drawn in the random selection were available for testing except six. One of these was sick, the other five were away on vacation. When it occurred that six of the selected teachers were not available for testing, the six teachers assigned to the next numbers appearing in the table of random numbers were chosen.

Description of Data Gathering Devices

1. Technical Agriculture Information

Technical agriculture competency was measured by administering tests of factual information in seven technical agriculture fields. These tests were in the areas of agronomy, dairy cattle, farm management, horticulture, poultry, soils, and swine. Three of the tests were constructed by the writer with guidance and supervision from an expert in the field at the University of Wisconsin.* Three tests were adapted from a battery of tests on animal enterprises by Ivan G. Fay, State Supervisor of the Veterans On-Farm Training Program in Wisconsin. This battery is entitled "Mastery Tests in Animal Husbandry", and is published by the Interstate, Danville, Illinois. Experts at the University of Wisconsin also assisted in the adaptation of these tests for use in this study.** Professor Edward J. Graul of the Soils Department of the University of Wisconsin prepared the test in soils. Competency in feeds and feeding was determined by

---

* Experts giving guidance and supervision to the construction of these tests were: Professor R. H. Andrew in agronomy; Professor Donald R. Mitchell in Farm Management; and Professor Malcolm N. Dana in horticulture.

**Experts assisting in this adaptation were: Professor James W. Crowley in dairy husbandry; Professor Robert H. Grummer in animal husbandry; and Professor James G. Halpin in poultry.

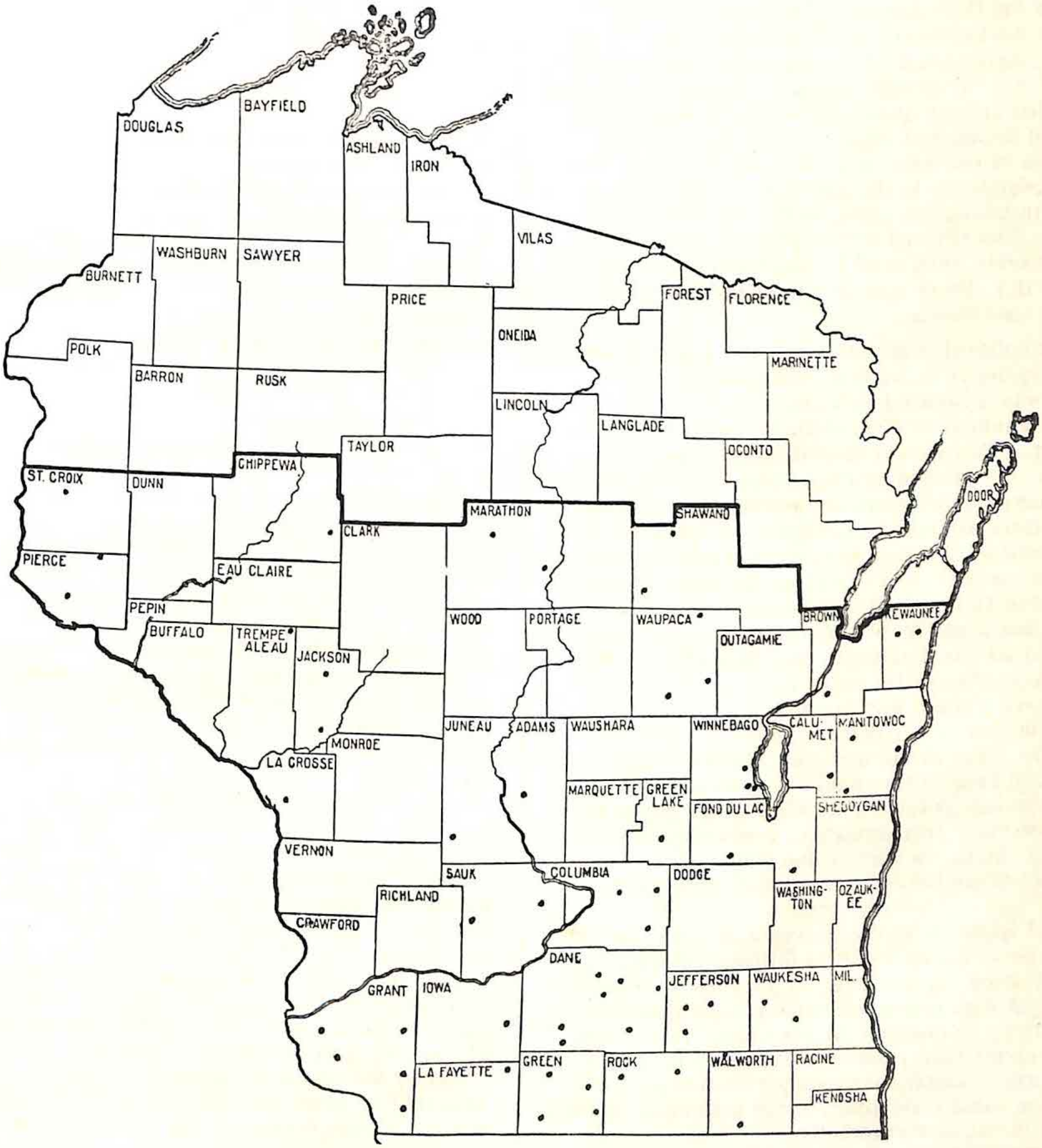
TABLE I

COMPARISON OF THE 16 VOCATIONAL AGRICULTURE TEACHERS ENROLLED IN SUMMER SCHOOL AT THE UNIVERSITY OF WISCONSIN WHO WERE DRAWN IN THE RANDOM SAMPLE SELECTION AND THE 14 TEACHERS WHO WERE NOT

Summer School Teachers Who Participated in Study	N	College Degrees Held	Reared on a Farm	Mean Age of Teachers (years)	Mean Years Experience in Teaching Agriculture	Mean Total High School Enrollment	Mean Total Vocational Agri. Dept. Enrollment
Drawn in random sample selection	16	16 with B. S. degree or equivalent	16	32. 69	7. 6	256. 25	89. 12
Not drawn in random sample selection	14	14 with B. S. degree or equivalent; 1 with M. S. degree	13	32. 86	8. 5	443. 71	81. 36

FIGURE 1

THE AREA FROM WHICH THE VOCATIONAL AGRICULTURE TEACHERS PARTICIPATING IN THIS STUDY WERE CHOSEN, AND THE DISTRIBUTION OF THE SCHOOLS IN WHICH THEY TAUGHT



Note: The heavy black line indicates the upper limits of the State from which participants in this study were chosen.

The black dots indicate the location of the 60 schools in which participants in this study taught.

the incorporation of questions on this topic into the animal enterprise tests.

In order to check the comprehensiveness of these tests, their coherence, and the approximate time required to take them, the original test drafts were administered to 11 upper classmen in the Department of Agricultural Education at the University of Wisconsin. (See Test drafts, Appendix A.)* In tabulating the responses to this "trial run" testing, it was discovered that certain questions had to be deleted or revised because of inappropriateness. Some sections of the tests were revised in order to give consistency to the question design, to add speed in taking the tests, and to facilitate tabulation. The revised technical agriculture information tests were used in this study. (See Appendix B.) There was no time limit for completing these tests.

2. Agricultural Manipulative Skills Information

Competency in agricultural manipulative skills was measured by a survey (see Appendix C) of manipulative farm skills completed by each subject. This survey of skills was adapted from a survey used by Rhoad (32) to measure the comprehensiveness of abilities in technical agriculture attained by prospective teachers of vocational agriculture in Ohio previous to their entrance into student teaching. This survey was limited to 15 farm enterprises or areas of farm skills that would be of general use and applicability in all rural areas in the state of Wisconsin. Also, the skills under each enterprise or area were limited to those that were considered highly desirable for farm people to perform proficiently. The areas included in this survey are: dairy and beef cattle, poultry, swine, corn, meadows and pastures, small grains, gardening, potatoes, fruit growing, concrete, electrification, farm machinery and power, erosion control, home beautification, and farm plumbing.

Each question in the survey was answered on a five-point rating scale as follows: 5) excellent, 4) good, 3) average, 2) poor, 1) no experience. A high score indicated a high degree of proficiency in performing the skill. Definition of the terms (see Appendix C) used in the rating scale were read by each subject so that all would have the same conception of the meaning of these terms in rating themselves.

A memorandum (see Appendix C) to each subject asked him to be objective in rating himself, to take his time (there was no time limit on this), and to rate himself on each item in the survey.

3. Professional Education Information

The measure of professional education competency of vocational agriculture teachers was accomplished by administering the Professional Information section of the 1951 edition of the National Teacher Examinations to all teachers participating in this study. Because of the highly confidential nature of these tests, and pursuant to directions from Mr. Arthur L. Benson, Director of Teacher Examinations of the Educational Testing Service, reproduction of these test questions cannot be included in this study. However, numerous references to the National Teacher Examinations can be found in the literature (20, 16, 33, 43), dating back to 1940 when they were first given.

The National Teacher Examinations are prepared and administered annually by the Educational Testing Service. They consist of a battery of tests designed to measure the professional background, mental ability, and general cultural knowledge of candidates for teaching positions (43). Only the Professional Information section of these tests was used in this study. Although this section yields a single score, it is composed of four parts. Each part of the 1951 edition of the Professional Information section of the test is timed; the total time allotted for taking this section is 75 minutes. There is a total of 105 questions in this section of the test. The four parts of this section of the test, the number of questions per part, and the time allotted per part are shown in Table II. All questions were of the objective type, consisting of short answers involving multiple-choice responses.

Since these examinations are revised annually, the reliabilities of the various tests vary slightly from year to year. The objective of the Educational Testing Service is to maintain minimum Kuder-Richardson reliability coefficients of .85 for the Professional Information section and other sections of the tests. Because of the detailed item analysis data available to the Educational Testing Service in constructing the tests, instances where they fail to meet this objective are rare. According to Mr. Benson, no part-score normative data are available for the 1951 edition of this test. In Appendix D of this study is a leaflet of the "Interpretation of Scores" for the 1951 nation-wide administration of the National Teacher Examinations. This leaflet, furnished by Mr. Benson, gives some interpretative data that includes information on the Professional Information section of the test that was used in this study.

4. Criteria of Success

The measure of teacher success was determined by ratings given to vocational agriculture teachers by state supervisors of vocational agriculture, principals, pupils, and farmers in the local school and community in which each teacher taught.

---

*All references to Appendices will be found in original thesis on file in Library, University of Wisconsin.

TABLE II

THE PARTS OF THE PROFESSIONAL INFORMATION SECTION OF THE 1951 EDITION OF THE NATIONAL TEACHER EXAMINATIONS, THE NUMBER OF QUESTIONS PER PART, AND THE TIME ALLOTTED FOR TAKING EACH PART

Parts of Professional Information Section of 1951 Edition of National Teacher Examination	No. of Questions per Part	Time Allotted for Taking Each Part
Part I, Education as a Social Institution	21	15 minutes
Part II, Child Development and Educational Psychology	28	20 minutes
Part III, Guidance and Measurement in Education	28	20 minutes
Part IV, General Principles and Methods of Teaching	28	20 minutes
Total	105	75 minutes

The state supervisor of vocational agriculture in Wisconsin and his staff cooperatively gave a rating to each teacher used in this study. They used a five-point rating scale as follows: 5) excellent, 4) good, 3) average, 2) below average, and 1) poor. A high score indicated a high rating; a low score indicated a low rating.

A short form of the Wisconsin Adaptation of the M-Blank (see Appendix E) was used by the principals to rate the vocational agriculture teachers in their schools. The rating under the item "General Evaluation" was used in this study. The five-point scale and the indications of the ratings were the same as those used for the supervisors.

An adaptation of a form used by Anderson (3) was used in this study to secure pupil ratings of vocational agriculture teachers. Mr. Anderson had adapted his pupil reaction to instruction form from a pupil evaluation device developed by Bryan. The all inclusive rating (see Appendix E) was the score used in this study. The five-point scale for this all inclusive rating was: 5) best fifth, 4) second fifth, 3) middle fifth, 2) fourth fifth, and 1) poorest fifth. A high score indicated a high rating; and a low score indicated a low rating.

A draft of a farmer rating scale (see Appendix F) was developed and administered to 15 farmers in a nearby rural community. The returned ratings of these farmers were analyzed, and the rating scale was revised (see Appendix G) for use in this study. The overall evaluation (item 6) was the score used in this investigation. The five-point scale and the indications of the rating were the same as those used for the supervisors.

Securing the Data

1. Competencies of Vocational Agriculture Teachers

Competencies of vocational agriculture teachers were measured by administering seven tests on factual information in technical agriculture, by administering a survey of manipulative skills in agriculture, and by administering a test on professional education information to the teachers participating in this study. Thirty of the participating teachers were enrolled in the 1953 summer school session at the University of Wisconsin. These several tests and the manipulative farm skills survey were administered to these teachers in a group on the campus of the University of Wisconsin in July, 1953. The remaining 30 teachers participating in this study were tested singly or in groups of two at the high schools in which they taught. The tests were administered to these teachers in July and August, 1953. Because of the security hazard involved in administering the highly confidential professional information examination, and in accordance with an agreement with Mr. Benson, Director of Teacher Examinations, Educational Testing Service, these tests were administered only in the presence of the writer, and under his continued supervision.

2. Success of Vocational Agriculture Teachers

Teacher success was determined from ratings given to the vocational agriculture teachers by state supervisors of vocational agriculture, principals, pupils, and farmers. A list of the teachers participating in this study was given to the state supervisor of vocational agriculture in September, 1953. In October, 1953, the state supervisor and his staff completed a cooperative rating of each teacher on the lists, and these ratings were used in the study.

In May, 1953, high school principals were sent a form letter (see Appendix E) asking them to rate their vocational agriculture teacher, and to have five officers of the local FFA rate him on forms enclosed for that purpose. These completed rating scales were returned in stamped self-addressed envelopes that were furnished by the writer.

Five farmer ratings were used for each vocational agriculture teacher participating in this study. The names and addresses of the farmers were secured from a general information sheet (see Appendix C) that each participating teacher completed at the time that he took the tests. Each teacher listed 10 farmers in his local school community. The five most cooperative farmers, according to the vocational agriculture teacher, were sent a form letter (see Appendix G) in August, 1953, asking them to rate the particular vocational agriculture teacher on a rating scale enclosed for the purpose. These completed rating scales were returned in stamped self-addressed envelopes that were furnished by the writer.

Treatment of the Data

As indicated earlier, seven different tests of technical agriculture were administered to teachers of vocational agriculture who participated in this study. These tests were in the areas of agronomy, dairy cattle, farm management, horticulture, poultry, soils, and swine. Also, a five-point rating scale survey of manipulative farm skills and a professional information test were administered. The manipulative farm skills survey was divided into fifteen farm enterprises or areas of farm skills. Finally, there was the rating of the efficiency of each teacher by four groups of raters as follows:

1. Rating of teacher by state supervisors.
2. Rating of teacher by principal on the Wiscon-

sin Adaptation M-Blank (short form).
3. Rating of teacher by pupils on an adaption of the Pupil Reaction to Instruction form used by Harold M. Anderson.
4. Rating of teacher by farmers on a Farmer Rating Scale.

Raw scores were used in all of the tabulations. All of the individual scores on each test and on each enterprise or area of manipulative farm skills survey were tabulated. After this was done, the following composites were made for each participating teacher:

1. Composite score made on the seven tests of technical agriculture. (Scores made on the four parts of the professional information test were originally recorded as a composite score.)

2. Composite score made on the manipulative farm skills survey.

3. Composite score made on the seven tests of technical agriculture and the manipulative farm skills survey.

4. Composite score made on the seven tests of technical agriculture and the professional information test.

5. Composite score made on the manipulative farm skills survey and the professional information test.

6. Composite score made on all testing devices; seven tests of technical agriculture manipulative farm skills survey, and the professional information tests.

Each rating on each participating teacher was tabulated. The principal ratings were secured as a single score with a possible high rating of five and a possible low rating of one for each teacher. The supervisory ratings were secured in a similar manner. The five pupil ratings were combined to make one score with a possible high rating of twenty-five and a possible low rating of five for each teacher. The five farmer ratings were treated the same as were the five pupil ratings. Then the principal and supervisory rating scores were adjusted to give them equal weight with the pupil and farmer rating scores. The adjustment was accomplished by multiplying five times the principal's rating score and five time the supervisor's rating score. This gave the principal and supervisory rating scores a possible high of twenty-five and a possible low of five for each teacher, thereby, equating these scores with the pupil and farmer rating scores. Using these adjusted scores, the following composites were made for each participating teacher:

1. Composite supervisory and principal rating score.
2. Composite supervisory and pupil rating score.
3. Composite supervisory and farmer rating score.
4. Composite principal and pupil rating score.
5. Composite principal and farmer rating score.
6. Composite pupil and farmer rating score.
7. Composite of the supervisory, principal, pupil, and farmer ratings.

Scores made on the several tests used in this study were intercorrelated. Scores made on the various enterprises or areas of the manipulative farm skills survey were intercorrelated, and intercorrelations were secured among efficiency rating scores. Intercorrelations were secured among the test scores and each efficiency rating, between each test score and all possible combinations of the ratings, as well as between a composite of the test scores and all possible combinations of the rating scores. This same procedure was followed in running intercorrelations among the manipulative farm skills survey and all the several ratings. Finally, intercorrelations were secured among composite scores made on the seven tests of technical agriculture, the manipulative farm skills survey, the professional educational information test, and the efficiency ratings.

From composite scores made on the tests in technical agriculture, the manipulative farm skills survey, and the professional educational tests, multiple regression techniques will be employed to develop equations for predicting the success of the vocational agriculture teacher. Composites of ratings will be used as the criteria of success. Finally, from the scores made on selected tests in technical agriculture, selected enterprises in the manipulative farm skills survey, and in the professional educational information test, multiple regression techniques will be employed to develop equations for predicting the success of the vocational agriculture teacher. Composites of ratings will be used as the criteria of success.

SECTION III

ANALYSIS OF DATA

IN THIS SECTION, the statistical treatment of the data will be presented. Occasional reference will be made to tables and to selected books for more detailed discussions of the data and methods employed in conducting this study.

The Pearson product moment coefficient of correlation (30) will be used in this study to indicate linear relationships existing between various sets of competency scores; between various sets of success rating scores; and between various competency and rating scores. *

Correlational data used in this study will be summarized with the use of observation, statistical significance of correlations (24), and regression equation techniques (12, 44). Correlations that are not significant at or below the 5% level of probability will not be considered statistically significant in this study.

This section will be divided into five parts. Part one will deal with the percentages of right answers made on the tests in technical agriculture, the percentage scores made on the manipulative farm skills survey, and the percentages of right answers on the test in professional education.

Part two will deal with the intercorrelations of success rating scores. Part three will be concerned with the intercorrelations of competency scores, while part four will deal with the correlations between competency scores and success rating scores. In part five the prediction of success will be considered.

Competency Scores and Success Rating Score Percentages

1. Test in Technical Agriculture

All tests in technical agriculture were designed and developed in similar manner. (See Section II.) They all deal with basic and fundamental factual material in their respective areas. Because of these similarities, the various tests of technical agriculture are assumed to have approximately equal difficulty and equal basic understanding requirements.

Table III indicates that the highest individual percentage score made on the test in technical agriculture was 96.3, made on the dairy cattle test. The mean percentage score on the dairy cattle test was 75.3. The only mean percentage score that was higher than this was made on the poultry test. The mean percentage score on the poultry test was 76.1.

The lowest individual percentage score, 32.9, was made on the soils test. The mean percentage score on the soils test was 63.5. The only mean percentage score that was lower than this was made on the farm management test. The mean percentage score on the farm management test was 60.7.

Higher individual percentage scores were made on the animal enterprise tests, including dairy cattle, poultry, and swine, than was made on any other technical agricultural test. Also, the animal enterprise test shows a higher low percentage score than any other single test except the agronomy test. The low percentage score on the agronomy test was 59.2. Except for agronomy and horticulture, with mean percentage scores of 75.1 and 75.0 respectively, the mean of the animal enterprise percentage scores was higher than it was for any other single test.

Farm management and soils are rather similar with respect to their mean, high, and low percentage scores. They have lower percentages of right answers in all three categories than any other single test. Soils has a higher mean and high score percentage right than farm management, while farm management has a higher low score percentage than soils.

Agronomy and horticulture are very similar with respect to their mean, high, and low percentage scores.

On the basis of the assumption that these tests in technical agriculture have approximately equal difficulty and equal basic understanding requirements, these percentage scores have the following indications: (1) that vocational agriculture teachers in Wisconsin, from whom this sample was chosen, have a better knowledge of basic facts in animal enterprises than they have in other farming enterprises investigated in this study; (2) that vocational agriculture teachers in Wisconsin, from whom this sample was chosen, have less knowledge of basic facts in soils and farm management than they have in other areas of farm endeavor investigated in this study, (3) that vocational agriculture teachers in Wisconsin, from whom this sample was chosen, have less knowledge of basic facts in agronomy and horticulture than they have in animal enterprises.

---

*The formula used in this computation was $r = \dfrac{N\Sigma XY - \Sigma X \cdot \Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}$ There were three assumptions made when computing these correlations. They were: (1) a linear relationship existed between each set of variables, (2) all the variables correlated had normal distributions, and (3) a line best fits its data when the sum of the squares of the deviations is a minimum. Assumptions one and two were made on the basis of actual inspection of the data. Assumption three is based on a principle generally accepted by mathematicians.

TABLE III

THE MEAN, HIGH, AND LOW SCORES MADE BY TEACHERS OF AGRICULTURE ON THE VARIOUS MEASURES EXPRESSED AS PERCENTAGES

	Mean Scores	High Score	Low Score
I. Tests in Technical Agriculture			
A. Agronomy	75.1	88.8	59.2
B. Dairy Cattle	75.3	96.3	57.3
C. Farm Management	60.7	73.5	42.6
D. Horticulture	75.0	88.9	57.1
E. Poultry	76.1	92.5	60.4
F. Soils	63.5	79.3	32.9
G. Swine	74.8	91.0	52.6
H. Composite of all technical agriculture tests	71.8	80.8	61.0
II. Tests in Professional Education	40.4	73.3	19.0
III. Manipulative Farm Skill Scores			
A. Dairy and Beef Cattle	70.0	93.3	41.3
B. Poultry	56.2	88.0	32.0
C. Swine	77.2	100.0	45.0
D. Corn	67.4	100.0	38.0
E. Meadow and Pasture	69.6	96.0	46.0
F. Small Grain	56.6	91.4	31.4
G. Gardening	52.3	85.0	25.0
H. Potatoes	70.2	88.0	20.0
I. Fruit Growing	67.8	100.0	44.0
J. Concrete	61.4	90.0	30.0
K. Electrification	61.1	100.0	34.3
L. Farm Machinery and Power	60.0	99.2	37.5
M. Erosion Control	57.4	100.0	20.0
N. Home Beautification	68.5	100.0	46.7
O. Farm Plumbing	66.5	100.0	35.0
P. Composite of Manipulative Farm Skills	62.0	91.0	40.6
IV. Composite of Individual Ratings of Teacher Success	85.0	100.0	61.7

TABLE IV

INTERCORRELATIONS OF INDIVIDUAL CRITERIA RATINGS OF VOCATIONAL AGRICULTURE TEACHER SUCCESS

	Principal	Farmer	Supervisory	Composite
Pupil	.443	.316	.320	.845
Principal	....	.307	.501	.640
Farmer	....	....	.488	.715
Supervisory	....	....	....	.644

However, their knowledge of basic facts in agronomy and horticulture seem greater than their knowledge of basic facts in soils and farm management.

The mean composite percentage of right answers made on all tests in technical agriculture was 71.8. There is no normative data available as yet on these tests, and no acceptable criterion was found for evaluating this score.

2. Test in Professional Education

Competency in professional education was measured by administering the 1951 edition of the professional information section of the National Teacher Examination to all teachers who participated in this study. (See Section II.)

Comparing the mean percentage score that the agriculture teachers made on this test (see Table III) with the mean scores and percentile ranks shown in Appendix D, it appears that the vocational agriculture teachers participating in this study are less well professionally informed than were the teachers taking this examination in 1951. However, this comparison must be qualified and considered in terms of the number of people taking the national administration of the test in 1951 as compared to the number of teachers used in this study.

This comparison can also be qualified in terms of the importance of this test to the individual teacher. Scores made on this test during the national administration of it appear very important to the person taking the test. Some school superintendents and boards of education require teaching applicants to submit scores made on the National Teacher Examinations to be evaluated as a part of their application. Also, the National Teacher Examinations are used as comprehensive examinations for undergraduates and as qualifying examinations for graduate students in some colleges. Because of the great importance of making high scores on the test, people taking the national administration of it in 1951 probably made special study and reviews of educational principles in preparation for taking the test. No such special preparation was made by the teachers participating in this study.

There is no data available for comparing scores made on the professional educational test by vocational agriculture teachers participating in this study with scores made on the same test by other vocational agriculture teachers. However, there is data available that seem to indicate that vocational teachers tend to make lower scores on professional educational test than do teachers of academic subjects. Table V of the leaflet in Appendix D indicates that industrial arts majors and home economics majors who took the 1946-50 national administrations of the test made lower mean scores than did most people who took the test during those years and who expressed teaching preferences other than in vocational subjects.

3. Manipulative Farm Skills

The manipulative farm skills score was secured with the use of a self-rating survey (see Section II), so the human factors inherent in rating oneself must be considered in analyzing these scores. Definition of terms used in the rating scale (see Appendix C) accompanied each survey so that all participating teachers would have similar understandings of term meanings and thereby reduce some of the subjectivity.

There does not appear to be any particular relationship among the mean, high, and low percentage scores made on the various enterprises in the manipulative farm skills survey. This would indicate a rather wide diversity of individual competencies in performing agricultural skills.

Swine, which had the lowest mean percentage score of the animal enterprise tests under test in technical agriculture above, has the highest mean percentage score of the animal enterprises under the manipulative farm skills survey. Dairy and beef cattle have the second highest mean percentage score of the animal enterprises under manipulative farm skills, and poultry, which had the highest mean percentage score of the three under technical agricultural tests, has the lowest mean percentage score under manipulative farm skills. Judging by tests and skills used in this study, this would indicate that there is very little relationship between one's knowledge of factual information in animal enterprises and his ability to perform animal enterprise skills. Table V, which shows intercorrelations of competency scores supports this analysis. This table shows that scores made on the animal enterprise tests of technical agriculture have almost no correlation with the scores made on the animal enterprise manipulative skills.

Intercorrelations of Rating Scores

The intercorrelations of individual criteria ratings can be seen in Table IV. Similar instructions were given each rater, and judging from the responses to the request for ratings, practically all raters cooperated readily. Although no reliability estimates of the raters used in the study were obtained, it is assumed that the ratings given each participating teacher was a valid and reliable opinion of the rater.

The supervisory and principal ratings are more closely correlated than any other two ratings. The supervisory and farmer ratings are the second most closely correlated ratings while the principal and farmer ratings are the lowest correlated ratings. All of the correlations are statistically significant. That is, the positive relationships that are indicated by the correlations are not due to chance alone.

Intercorrelations of Competency Scores

1. Test in Technical Agriculture

Among the technical agricultural tests are a considerable number of statistically significant correlations; some are higher than others. Table V shows that the highest correlations are among the animal enterprise tests. Correlations of .735 between the swine and dairy cattle tests, .623 between the poultry and dairy cattle tests, and .576 between the swine and poultry tests, are statistically significant at the .1% level. No other correlation is statistically significant at such a high level.

Among the low correlations in the technical agricultural test scores is noted the correlation of .084 between poultry and horticulture, .118 between the dairy cattle and agronomy tests, and .169 between the swine and soils tests.

There is a consistently positive correlation between all of the technical agricultural test scores. This indicates that agriculture instructors who ranked high on one test tended to rank high on other tests, and agriculture instructors who ranked low on one test tended to rank low on other tests.

2. Manipulative Farm Skills

There are quite a number of negative correlations between the technical agricultural test scores and the manipulative farm skills survey scores. Table V shows that only one of the correlations between technical agricultural test scores and manipulative farm skills survey scores is statistically significant. This one is between the swine technical agricultural test and the swine manipulative farm skills. An example of a negative correlation between a technical agricultural test score and a manipulative farm skills score is the correlation of -.125 between the dairy and beef cattle manipulative skill and the agronomy test score. An example of a positive correlation between a technical agricultural test score and a manipulative farm skills score is the correlation of .057 between the poultry manipulative skill score and the agronomy test score. The consistently low correlations between the manipulative farm skill scores and the technical agricultural test score indicate that there is very little relationship between a vocational agriculture teacher's knowledge of factual information in agriculture and his own judgment regarding his ability to perform agricultural manipulative skills.

Intercorrelations among the manipulative farm skill scores give consistently positive correlations. Practically all of these manipulative farm skill score inter-correlations are statistically significant. All intercorrelations of the animal enterprise manipulative farm skill scores are statistically significant at the .1% level. Most intercorrelations of the mechanical and farmstead type of manipulative farm skill scores are statistically significant at the .1% level.

The only correlations among the manipulative farm skill scores that are not statistically significant are between gardening and poultry, concrete and gardening, erosion control and gardening, erosion control and potatoes, home beautification and potatoes, home beautification and concrete, home beautification and electrification, and home beautification and swine.

The consistently positive correlations among the manipulative farm skill scores indicate that agriculture teachers who ranked themselves high in performing one manipulative skill tended to rank themselves high in performing other manipulative skills. Also, agriculture teachers who ranked themselves low in performing one manipulative skill tended to rank themselves low in performing other manipulative skills.

3. Professional Educational Test

As shown in Table V, the professional education test scores correlate positively with all of the technical agriculture test scores. The correlation between scores made on the professional education test and the farm management test is statistically significant at the .1% level. There are statistically significant correlations at the .1% level between the professional education and the dairy cattle test scores; between the professional education and the poultry test scores; and between professional education and the swine test scores. The correlation between professional education and the horticulture test score is statistically significant at the 5% level, while the correlations between professional education and the soils test score and between professional education and the agronomy test score are not statistically significant.

All correlations between professional education test scores and manipulative farm skill scores are negative except one, and only one is statistically significant. This gives the indication that an agriculture teacher who ranked high on the professional education test tended to rank high on the technical agriculture test. Also, an agriculture teacher who ranked low on the professional education test tended to rank low on the technical agriculture test. It indicates that there is very little relationship between a vocational agriculture teacher's knowledge of professional education information and his own judgment regarding his ability to perform manipulative farm skills.

TABLE V

THE INTERCORRELATIONS OF AGE, YEARS OF TEACHING EXPERIENCE, TECHNICAL AGRICULTURE TEST SCORES, MANIPULATIVE FARM SKILL SCORES, PROFESSIONAL EDUCATION TEST SCORES, AND COMBINATIONS OF COMPOSITES OF COMPETENCY SCORES

		Tests in Technical Agriculture								Manipulative Farm Skill			
	Experience Teaching	Agronomy	Dairy Cattle	Farm Management	Horticulture	Poultry	Soils	Swine	Composite	Dairy and Beef Cattle	Poultry	Swine	Corn
Age	909	031	114	115	257	140	019	086	145	247	208	066	277
Experience in Teaching	....	073	179	157	273_1	228	042	218	225	243	240	096	153
Tests in Technical Agriculture													
Agronomy	....	....	118	281_1	359_2	265_1	236	250	513_3	-125	057	-057	-134
Dairy Cattle	....	....	....	318_1	263_1	623_3	255	735_3	752_3	038	028	053	-148
Farm Management	....	....	....	....	230	264_1	430_2	340_1	540_3	-011	119	047	089
Horticulture	....	....	....	....	....	084	319_1	379_2	485	015	-101	105	-101
Poultry	....	....	....	....	....	....	290_1	576_3	723_3	010	052	-065	-151
Soils	....	....	....	....	....	....	....	169	580_3	-016	018	-133	-116
Swine	....	....	....	....	....	....	....	....	774_3	128	183	312_1	-062
Composite	....	....	....	....	....	....	....	....	....	005	090	077	-133
Manipulative Farm Skills													
Dairy and Beef Cattle	....	....	....	....	....	....	....	....	....	....	633_3	651_3	500
Poultry	....	....	....	....	....	....	....	....	....	....	....	569_3	559
Swine	....	....	....	....	....	....	....	....	....	....	....	....	492
Corn	....	....	....	....	....	....	....	....	....	....	....	....	....
Meadow and Pasture	....	....	....	....	....	....	....	....	....	....	....	....	....
Small Grain	....	....	....	....	....	....	....	....	....	....	....	....	....
Gardening	....	....	....	....	....	....	....	....	....	....	....	....	....
Potatoes	....	....	....	....	....	....	....	....	....	....	....	....	....
Fruit Growing	....	....	....	....	....	....	....	....	....	....	....	....	....
Concrete	....	....	....	....	....	....	....	....	....	....	....	....	....
Electrification	....	....	....	....	....	....	....	....	....	....	....	....	....
Farm Machinery and Power	....	....	....	....	....	....	....	....	....	....	....	....	....
Erosion Control	....	....	....	....	....	....	....	....	....	....	....	....	....
Home Beautification	....	....	....	....	....	....	....	....	....	....	....	....	....
Farm Plumbing	....	....	....	....	....	....	....	....	....	....	....	....	....
Composite	....	....	....	....	....	....	....	....	....	....	....	....	....
Professional Education Test	....	....	....	....	....	....	....	....	....	....	....	....	....
Technical Agriculture and Professional Education	....	....	....	....	....	....	....	....	....	....	....	....	....
Professional Education and Manipulative Skills	....	....	....	....	....	....	....	....	....	....	....	....	....
Technical Agriculture and Manipulative Skills	....	....	....	....	....	....	....	....	....	....	....	....	....

Decimal points have been omitted.

$_1$Significant at 5% level
$_2$Significant at 1% level
$_3$Significant at .1% level

Manipulative Farm Skills											Combinations of Composites of Scores			
Gardening	Potatoes	Fruit Growing	Concrete	Electrification	Farm Machinery and Power	Erosion Control	Home Beautification	Farm Plumbing	Composite	Professional Education Test	Technical Agriculture and Professional Ed.	Professional Ed. and Manipulative Farm Skills	Technical Ag. and Manipulative Farm Skills	Total Composite of Competencies
465_3	151	276_1	078	228	365_2	093	362_2	289_1	384_2	-138	070	372	417	400
407_2	081	275_1	054	235	342_2	078	409_2	313_1	353_2	-137	134	341	406	390
-208	-049	-153	-018	045	-145	067	012	-161	-130	234	490	-106	-001	024
-164	-084	-108	-010	-200	-165	-176	-006	-060	-159	352_2	720	-123	030	067
-086	004	-047	128	-051	212	025	060	090	167	466_3	588	217	300	347
031	-038	060	096	198	080	222	112	037	092	271_1	479	121	212	239
-098	-031	-210	-097	-232	-130	-211	-022	-175	-153	344_2	694	-117	029	065
-009	-055	-230	-088	-064	145	-107	023	-072	049	234	543	074	093	216
-035	-027	010	109	017	-055	-058	075	014	012	373_2	745	027	181	219
-124	-065	-199	-018	-106	-073	-148	-026	-104	-089	459_3	954	-041	161	208
384_2	357_2	418_2	287_1	309_1	430_2	488_3	311_1	388_2	601_3	-179	-056	586	597	575
207	588_3	381_2	427_2	491_3	526_3	370_2	443_3	372_2	690_3	-062	052	688	706	696
355_2	337_1	317_1	390_2	408_2	348_2	349_2	255	400_2	533_3	-150	012	521	548	529
587_3	504_3	451_3	523_3	489_3	700_3	352_2	340_1	500_3	809_3	-101	-141	803	768	753
583_3	511_3	437_2	294_1	339_1	458_3	298_1	413_2	424_2	623_3	-087	-102	618	595	582
469_3	515_3	393_2	319_1	260_1	492_3	388_2	332_1	337_1	109	-158	-161	581	554	535
....	475_3	485_3	252	363_2	478_3	232	626_3	445_3	624_3	-301_1	-201	596	587	553
....	....	373_2	462_3	357_2	411_2	170	225	286_1	544_3	-053	-070	542	522	514
....	....	....	299_1	367_2	349_2	368_2	590_3	393_2	509_3	-067	-182	505	455	445
....	....	....	....	521_3	487_3	360_2	132	523_3	573_3	057	004	583	563	566
....	....	....	....	....	488_3	325_1	232	621_3	604_3	-034	113	599	572	560
....	....	....	....	....	....	417_2	349_2	567_3	946_3	-061	-079	445	919	906
....	....	....	....	....	....	....	265_1	324_1	515_3	-247	-201	493	474	446
....	....	....	....	....	....	....	....	353_2	473_3	-096	-053	466	462	450
....	....	....	....	....	....	....	....	....	668_3	-038	-096	668	636	628
....	....	....	....	....	....	....	....	....	....	-117	-111	994	969	951
....	....	....	....	....	....	....	....	....	....	....	703	-011	-001	103
....	....	....	....	....	....	....	....	....	....	....	....	-037	128	201
....	....	....	....	....	....	....	....	....	....	....	....	....	975	969
....	....	....	....	....	....	....	....	....	....	....	....	....	....	995

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated
Acc. No

The correlations of the composite combinations of scores, shown in Table V, are without statistical meaning, and therefore the statistical significance of them was not computed. These composite scores are often correlated with themselves which automatically raises the level of correlation. Where there are instances of a composite that has a high possible score being correlated with itself and a composite that has a low possible score, the resulting correlation is disproportionately high. An example of this can be seen between the composite scores of the professional education test and the manipulative farm skill survey being correlated with the manipulative farm skills composite. The correlation that results is .994. However, the manipulative farm skills score that is being correlated with itself has a possible high of 1065, while the professional education test has a possible high score of only 105. The scores actually made on these instruments (manipulative farm skills survey and professional education test) makes this improper balance even worse. For example, the mean score made on the manipulative farm skills survey was 660 which is 62 percent of the highest possible score, while the mean score of right answers made on the professional educational test was 42 which is 40 percent of the highest possible score.

Correlations of Competency Scores and Success Rating Scores

1. Individual Ratings

The correlations between competency scores and individual rating scores are presented in Table VI. Inspection of this table reveals that there are more statistically significant correlations between farmer rating scores and competency scores and between pupil rating scores and competency scores than there are between any other individual rating scores and competency scores. There is a statistically significant correlation at the 5% level between the farmer rating score and the poultry manipulative farm skills score. There are statistically significant correlations at the 1% level between the farmer rating score and the farm management test, between the farmer rating score and the poultry test, and between the farmer rating score and the composite of technical agriculture test scores.

All correlations between pupil rating scores and competency scores are statistically significant at the 5% level. These correlations exist between the pupil rating scores and scores made on the following tests: farm management, horticulture, soils, and the composite of technical agriculture test scores.

There is a 5% level of statistical significance between the principal rating scores and the farm management test scores, between the principal rating scores and the poultry test scores, and between the principal rating scores and the composite of technical agriculture test scores.

There is a 5% level of statistical significance between the supervisory rating score and the farm management test score. The supervisory rating score has no other statistically significant correlation. However, it correlates higher with the composite of the manipulative farm skills score than does any other individual rating score. The farmer rating score correlates higher with the composite scores of the technical agriculture test and the professional education test than does any other individual rating score.

The composite of the individual rating scores have a statistically significant correlation with a greater number of the competency scores than does any of the rating scores taken individually. Significantly, the composite of the individual rating scores correlates at the 1% level of statistical significance with the composite of technical agriculture test scores. It does not have a statistically significant correlation with either the composite manipulative skills score or the professional education test.

The correlations between each of the individual rating scores and each of the competency composite scores are consistently positive. Each individual rating score has a statistically significant correlation with the farm management test score.

The correlations between individual rating scores and the technical agriculture test scores are generally higher than they are between the individual rating scores and either the manipulative farm skills score or the professional education test score. The correlations between the individual rating scores and the professional education test scores are generally higher than they are between the individual rating scores and the manipulative farm skill scores. Judging by ratings, test, and skills used in this study, the above would indicate that there is a closer relationship between the success of a vocational agriculture teacher, as measured by individual ratings, and his knowledge of basic facts in technical agriculture than there is between his success, measured by the same criteria, and his knowledge of professional education or his judgment of his ability to perform manipulative farm skills. There seems to be a closer relationship between his success, as measured by the above criteria, and his knowledge of professional education than there is between his success and his judgment of his ability to perform manipulative farm skills.

2. Combinations of Ratings

The correlations between competency scores and combinations of rating scores are presented in Table VI. This table reveals that the great majority of statistically significant correlations are found between the combinations of rating scores and the technical agriculture test scores. Combinations of rating scores have statistically significant correlations with only three of the manipulative farm skills enterprise scores. Three of the combinations of rating scores have statistically significant correlations with the professional education test score.

Five of the combinations of rating scores have statistically significant correlations with six of the competency score columns. These combinations of rating scores are: pupil-farmer, pupil-farmer-supervisory, principal-farmer-supervisory, principal-farmer, and the composite of combination rating scores.

Four of the combinations of rating scores have statistically significant correlations with five of the competency scores. These combinations of rating scores are: pupil-principal, pupil-supervisory, pupil-principal-farmer, and pupil-principal-supervisory rating scores. Principal-supervisory and farmer-supervisory rating scores have statistically significant correlations with four of the competency scores.

The composite of combinations of rating scores correlates at the 1% level of statistical significance with the composite of technical agriculture test scores, and it correlates at the 5% level of statistical significance with the professional education test scores. The correlation between the composite of combinations of rating scores and the composite score of manipulative farm skills is positive but not statistically significant.

The Prediction of Success

Table VII presents correlations used in rejecting or accepting the first three null hypotheses postulated in this study. The correlations between the composite of technical agriculture test scores and the composite of individual rating scores, and between the composite of technical agriculture test scores and the composite of combinations of rating scores are statistically significant at the 1% level. This means that there is a statistically significant relationship between the technical agriculture test scores and the success of vocational agriculture teachers, success being measured by composite of rating scores. On the basis of these correlations being statistically significant from zero, the first null hypothesis postulated in this study was rejected.

Having established that there is a statistically significant relationship between a vocational agriculture teacher's competency in technical agriculture and his success as a teacher of vocational agriculture, it was decided to determine the predictive value of the coefficient of correlation. This was done in part by computing the standard error of estimate. (See Table VII.) The formula (12, 47) used in this computation was $Sy = \sigma y\sqrt{1 - r^2}$, where σy = standard deviation of the predicted scores (composite of individual rating scores), and r = the coefficient of correlation between the two variables. Substituting into the formula give $Sy = 4.64\sqrt{1-(.387)^2}$. The standard error of estimate was found to be 3.94 for the correlation between the composite of technical agriculture test scores and the composite of individual rating scores. Substituting into the above formula for the correlation between the composite of technical agriculture test scores and the combination of composites of rating scores, we get $Sy = 9.57\sqrt{1 - (.362)^2}$, which equals 8.32. So, the error that might be expected in predicting individual scores where r = .362 is more than twice as great as the error that might be expected in predicting individual scores where r = .387.

More specifically, the ratio of the residual variance [or the value 1 - r (12, 47)] to the total variance is found to be .85 for r = .387 and .87 for r = .362. (See Table VII.) This means that only 15 percent of the total variance of scores is accounted for by the first coefficient of correlation, and 13 percent of the total variance of scores is accounted for by the second coefficient of correlation. Thus, while both coefficients of correlation may have some predictive value for groups of vocational agriculture teachers, they have very limited value for predicting individual scores.

As may be seen in Table VII, the correlations between the composite of manipulative farm skill scores and the composite of individual rating scores, and between the composite of manipulative farm skill scores and the composite of combinations of rating scores are not statistically significant. This means that there is no statistically significant relationship between the manipulative farm skill scores and the success of vocational agriculture teachers, success being measured by composites of rating scores. On the basis of these correlations being statistically insignificant from zero, the second null hypothesis postulated in this study was accepted.

Using the same reasoning, methods, and procedures as those above, the standard error of estimate and the ratio of residual variance to the total variance were calculated for r = .131 and r = .146. Results of these calculations appear in Table VII. Neither correlation has any value for predicting group scores nor for predicting individual scores.

Again referring to Table VII, it may be seen

TABLE VI

THE CORRELATIONS OF CRITERIA RATING SCORES WITH AGE, YEARS OF TEACHING EXPERIENCE, TECHNICAL AGRICULTURE TEST SCORES, MANIPULATIVE FARM SKILL SCORES, PROFESSIONAL EDUCATION TEST SCORES, AND COMBINATIONS OF COMPOSITES OF COMPETENCY

			Technical Agriculture Test Scores									Manipulative Farm Sk		
Individual Raters	Age	Experience Teaching	Agronomy Test	Dairy Cattle Test	Farm Management Test	Horticulture Test	Poultry Test	Soils Test	Swine Test	Composite of Technical Agriculture Test Scores	Dairy and Beef Cattle	Poultry	Swine
Pupil	015	016	078	091	317_1	263_1	114	279_1	222	263_1	243	178	16
Principal	-041	013	153	154	309_1	068	335_1	239	244	301_1	114	014	20
Farmer	000	038	072	257	358_2	129	436_2	199	236	364_2	124	283_1	07
Supervisory	149	164	-001	245	258_1	069	209	180	216	219	253	134	20
Composite of Individual Raters	099	056	100	224	425_2	230	331_1	315_1	303_1	387_2	256	241	19
Combinations of Raters													
Pupil-Principal	-021	017	143	150	366_2	173	286_1	299_1	-275_1	334_1	195	095	22
Pupil-Farmer	011	030	093	191	407_2	256	295_1	302_1	278_1	370_2	239	268_1	16
Pupil-Supervisory	117	129	037	224	342_2	177	170	267_1	266_1	287_1	303_1	184	25
Pupil-Principal-Farmer	-057	-012	094	162	364_2	102	367_2	288_1	265_1	334_1	219	193	25
Pupil-Principal-Supervisory	059	091	094	219	368_2	148	291_1	286_1	287_1	328_1	253	128	24
Pupil-Farmer-Supervisory	098	118	050	256	382_2	183	291_1	276_1	285_1	338_1	287_1	229	21
Principal-Farmer	-033	025	152	223	388_2	104	433_2	271_1	288_1	383_2	140	118	19
Principal-Supervisory	066	105	085	232	326_1	079	311_1	241	265_1	298_1	215	087	25
Principal-Farmer-Supervisory	058	100	090	261_1	366_2	098	371_2	256	285_1	343_2	216	140	22
Farmer-Supervisory	118	142	024	280_1	325_1	098	313_1	210	251	297_1	242	202	18
Composite of Combinations of Raters	053	089	098	244	397_2	157	343_2	294_1	302_1	362_2	250	168	25

Decimal points have been omitted.

$_1$Significant at the 5% level.
$_2$Significant at the 1% level.

Manipulative Farm Skill Scores												Combinations of Composites of Scores			
Small Grain	Gardening	Potatoes	Fruit Growing	Concrete	Electrification	Farm Machinery and Power	Erosion Control	Home Beautification	Farm Plumbing	Composite of Manipulative Farm Skill Scores	Professional Education Test	Technical Agricultural and Professional Education	Professional Education and Manipulative Farm Skills	Technical Agricultural and Manipulative Farm Skills	Total Composite of Competencies
-026	021	-116	-032	-061	-049	046	104	-052	-170	065	106	246	077	130	140
101	-180	134	-018	069	-096	098	-033	-051	-100	082	240	321	109	157	181
-084	063	187	066	101	-050	123	-020	030	-068	128	248	375	155	214	242
011	-074	059	030	193	-081	167	049	092	057	160	202	243	183	213	233
-029	-007	038	009	050	-079	122	056	-011	-133	131	236	389	157	226	250
057	-114	037	-028	018	-090	090	024	-060	-151	088	217	340	112	171	192
-060	046	006	008	001	-060	093	067	-023	-158	110	199	363	132	201	221
-004	-044	-012	007	113	-083	145	086	043	-040	149	200	297	172	220	239
007	-060	074	-089	068	-048	124	015	-167	-089	126	234	346	152	208	232
043	-112	053	-004	106	-099	141	040	005	-072	137	242	344	163	217	241
-026	-020	039	023	121	-083	154	067	004	-051	159	234	349	185	242	265
051	-124	180	010	094	-097	126	-038	-030	-108	116	290_1	404	147	210	239
063	-144	110	008	153	-102	154	008	027	-022	141	255	324	170	215	240
036	-112	138	021	157	-100	163	003	030	-034	153	279_1	368	183	237	265
-019	-037	110	046	186	-081	173	032	083	022	170	244	319	197	242	266
023	-088	083	009	114	-098	149	032	011	-077	146	264_1	379	156	235	262

1.9.70
J302

TABLE VII

THE CORRELATIONS OF COMPETENCY VARIABLES WITH SUCCESS CRITERIA, THEIR STANDARD ERROR OF ESTIMATE, AND RESIDUAL VARIANCE

	Composite of Technical Agriculture Test Scores			Composite of Manipulative Farm Skill Survey Scores			Professional Education Test Scores		
	r	Sy	R-V[1]	r	Sy	R-V	r	Sy	R-V
Composite of individual rating scores	.387[2]	3.94	.85	.131	4.56	.98	.236	4.38	.94
Combination of composites of rating scores	.362[2]	8.32	.87	.146	9.36	.97	.264[3]	8.89	.93

[1] Residual Variance
[2] Significant at 1% level.
[3] Significant at 5% level.

TABLE VIII

CORRELATION MATRIX FOR SOLVING REGRESSION EQUATION NUMBER ONE

	X_1	X_2	X_3	X_4
X_1	1.000	.387	.131	.236
X_2	.387	1.000	-.089	.459
X_3	.131	-.089	1.000	-.117
X_4	.236	.459	-.117	1.000

TABLE IX

CORRELATION MATRIX FOR SOLVING REGRESSION EQUATION NUMBER TWO

	X_1	X_2	X_3	X_4
X_1	1.000	.362	.146	.264
X_2	.362	1.000	-.089	.459
X_3	.146	-.089	1.000	-.117
X_4	.264	.459	-.117	1.000

that the correlation between the professional education test and the composite of individual rating scores is not statistically significant. In order to be statistically significant at the 5% level, this correlation would have to be at least .258. The correlation of .264 between the professional education test and the combination of composites of rating scores, however, is statistically significant at the 5% level. This indicates that there is no statistically significant relationship between the professional education test score and the success of vocational agriculture teachers when success is measured by a composite of individual ratings. It also indicates, however, that there is a statistically significant relationship between the professional education test score and the success of vocational agriculture teachers when success is measured by a composite of combinations of rating scores. On the basis of this finding, the third null hypothesis postulated in this study is neither accepted nor rejected. It is left open for question and further study.

The high percentage of residual variance found for both correlations is interpreted to mean that neither correlation has much value in predicting individual scores.

It was decided that in order to check the fourth null hypothesis postulated in Section I, two regression equations would be calculated. In the first regression equation the technical agriculture test composite score, the manipulative farm skills survey score, and the professional education test score were used to predict the composite of individual rating scores. In the second the same competency scores listed above were used to predict the composite combinations of rating scores.

For solving the first equation mentioned above, the following symbols were assigned to the variables: X_1, composite of individual ratings; X_2, composite of technical agriculture test scores; X_3, composite of manipulative farm skills survey scores; and X_4, professional education test scores. Then their intercorrelations were calculated and placed in a correlation matrix. (See Table VIII.) This matrix was then solved, using Aitkin Modified Method of Pivotal Condensation (44, 12). Using the multiple correlation formula (12), $R_{1.234} = \sqrt{B_2 r_{12} + B_3 r_{13} + B_4 r_{14}}$ the beta coefficients were found to be $R_{1.234} = \sqrt{(.361)(.387) + (.173)(.131) + (.091)(.236)}$. The regression equation for all four variables yields a multiple correlation of .429.

By using the formula $t = \frac{r\sqrt{f}}{\sqrt{1-R^2}}$ and the t table in Statistical Methods in Research by Johnson (24), the multiple correlation of .429 was found to be statistically significant at the 1% level. This means that there is a statistically significant relationship between the combination of competencies and success of vocational agriculture teachers, success being measured by a composite of individual rating scores. On the basis of this correlation being statistically significant from zero, the fourth null hypothesis postulated in this study was rejected, pending the findings of the second regression equation yet to be calculated.

After concluding that the competencies have some relationship to success as measured by the composite of individual rating scores, the next step was to determine the predictive value of the prediction equation. The formula $S_y = \sigma_y\sqrt{1 - R^2}$ was used to calculate the standard error of estimate. Substituting into the formula gives $S_y = 4.64\sqrt{1 - (.429)^2}$, which equals 4.19. An example of the use of this standard error of estimate score is given by taking one actual predicted score of 52.42. Using this actual predicted score of 52.42, one can predict that 95.45 times out of 100 the true score would be expected to fall between 44.04 and 60.80.

The ratio of the residual variance to the total variance for this prediction equation is .82. This means that only 18 percent of the total variance of scores is accounted for by this particular regression equation. So, while the regression equation may have some predictive value for groups of vocational agriculture teachers, it has limited value for predicting individual scores.

In order to check the reliability of the composite of the technical agriculture test score, composite of manipulative farm skill survey score, and the professional education test score to predict vocational agriculture teacher's success as measured by the composite of combinations of rating scores, a second regression equation was calculated. The following symbols were assigned to the variables that were used in this equation: X_1, composite of combinations of rating scores; X_2, composite of technical agriculture test scores; X_3, composite of manipulative farm skills survey scores; and X_4, professional education test scores. The matrix formed by the intercorrelation of these variables is presented in Table IX.

Using the same methods and procedure as those used in the first equation, it was found that the four variables used in this regression equation yields a multiple correlation of .423. This correlation is statistically significant at the 1% level. Therefore, there is a statistically significant relationship between the combination of competencies used in the regression equation, and the success of vocational agriculture teachers, success being measured by a composite of combinations of rating scores. Since this correlation is statistically significant from zero, it supports the finding of the first regresion equation and the fourth null hypothesis postulated in this study was rejected.

The standard error of estimate for the second regression equation is 8.67. The ratio of the residual variance to the total variance is .82, which is the same ratio as that found for the first equation. The conclusion is drawn that while this regression equation may have some predictive value for groups of vocational agriculture teachers, it might be considered as having limited value for predicting individual scores.

In the first analysis of the data, the correlations of various enterprises under the technical agriculture tests and of various enterprises under the manipulative farm skills survey seemed to be sufficiently high to warrant further study in search of more reliable regression equations for predictive purposes. It was then decided to calculate four additional regression equations using the same methods and procedures as previously reported. These additional regression equations will make the third, fourth, fifth, and sixth regression equations used in this study.

The third regression equation involved the use of variables X_2, the farm management test score; X_3, the dairy and beef cattle manipulative farm skills score; and X_4, the professional education test score to predict X_1, the composite of individual rating scores. The matrix formed by the intercorrelation of these variables is presented in Table X. The regression equation for these four variables yields a multiple correlation of .508. This correlation is higher than either of the previous multiple correlations. It is statistically significant at the .1% level. The standard error of estimate for this equation is 4.00. This standard error of estimate is smaller than the standard error of estimate for either of the above two equations, indicating that the range within which one might expect a predicted score to fall is not as great as the range within which one might expect a predicted score to fall for the first and second regression equations used in this study.

The ratio of the residual variance to the total variance is .74. This means that 26 percent of the total variance of scores is accounted for by this particular regression equation. So, while this regression equation has more value in predicting individual scores than either of the above equations, it still does not account for enough of the total variance to be of great value for that purpose.

The fourth regression equation involved the use of variables X_2, the farm management test score; X_3, the dairy and beef cattle manipulative farm skills score; and X_4, the professional education test score to predict X_1, the composite of combinations of rating scores. The matrix formed by the intercorrelation of these variables is presented in Table XI. The multiple correlation of this regression equation is .492. This multiple correlation is statistically significant at the .1% level. The standard error of estimate for this equation is 8.34, and the ratio of the residual variance to the total variance is .76. As was true with the previous regression equations, this one has no great value for predicting individual scores.

The fifth regression equation involved the scores made on two of the technical agriculture test enterprises, two of the manipulative farm skill survey enterprises, the professional educational test, and the composite of individual rating scores. The variables and the symbols assigned to them are: X_2, the farm management test score; X_3, the poultry test score; X_4, the dairy and beef cattle manipulative farm skills score; X_5, the poultry manipulative farm skills score; X_6, the professional education test score; and X_1, the composite of individual rating scores. Variables X_2 through X_6 were employed to predict X_1. The matrix formed by the intercorrelation of these variables is presented in Table XII. The multiple correlation for this regression equation is .486, making it statistically significant at the .1% level. The standard error of estimate is 4.06 and the ratio of the residual variance to the total variance is .76. This is the same ratio as was found in the preceding equation, so the same conclusion can be drawn from it. It should be noted that by adding the two factors to the variables used in regression equation number three the multiple correlation is lowered and not raised.

The sixth and last regression equation calculated involved the use of variables X_2, the farm management test score; X_3, the poultry test score; X_4, the dairy and beef cattle manipulative farm skills score; X_5, the meadow and pasture manipulative farm skills score; and X_6, the professional education test score to predict X_1, the composite of combinations of rating scores. The matrix formed by the intercorrelations of these variables is presented in Table XIII. This regression equation has a multiple correlation of .468, which is statistically significant at the .1% level. The standard error of estimate is 8.46 and the ratio of the residual variance to the total variance is .78. This means that only 22 percent of the total variance of scores is accounted for by this regression equation and therefore, it has no great value in predicting individual scores. It is worthy of note that by adding the two factors to the variables used in regression equation number four the multiple correlation is lowered.

Table XIV presents the multiple correlations, levels of statistical significance, standard error of estimates, and residual variances for the six regression equations calculated in this study. In comparing Table XIV with Table VII, it is noted

TABLE X

CORRELATION MATRIX FOR SOLVING REGRESSION EQUATION NUMBER THREE

	X_1	X_2	X_3	X_4
X_1	1.000	.425	.256	.236
X_2	.425	1.000	- .011	.466
X_3	.256	- .011	1.000	- .179
X_4	.236	.466	- .179	1.000

TABLE XI

CORRELATION MATRIX FOR SOLVING REGRESSION EQUATION NUMBER FOUR

	X_1	X_2	X_3	X_4
X_1	1.000	.397	.250	.264
X_2	.397	1.000	- .011	.466
X_3	.250	- .011	1.000	- .179
X_4	.264	.466	- .179	1.000

TABLE XII

CORRELATION MATRIX FOR SOLVING REGRESSION EQUATION NUMBER FIVE

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1.000	.425	.331	.256	.241	.236
X_2	.425	1.000	.264	- .011	.119	.466
X_3	.331	.264	1.000	.010	.052	.344
X_4	.256	- .011	.010	1.000	.633	- .179
X_5	.241	.119	.052	.633	1.000	- .062
X_6	.236	.466	.344	- .179	- .062	1.000

TABLE XIII

CORRELATION MATRIX FOR SOLVING REGRESSION EQUATION NUMBER SIX

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1.000	.397	.343	.250	.234	.264
X_2	.397	1.000	.264	- .011	.200	.466
X_3	.343	.264	1.000	.010	- .094	.344
X_4	.250	- .011	.010	1.000	.427	- .179
X_5	.234	.200	- .094	.427	1.000	- .087
X_6	.264	.466	.344	- .179	- .087	1.000

TABLE XIV

MULTIPLE CORRELATIONS, LEVELS OF SIGNIFICANCE, STANDARD ERROR OF ESTIMATES, AND RESIDUAL VARIANCES FOR SIX REGRESSION EQUATIONS

Regression Equation*	Multiple Correlations	Levels of Significance	Standard Error of Estimates	Residual Variance
1	.429	1%	4.19	.82
2	.423	1%	8.67	.82
3	.508	.1%	4.00	.74
4	.492	.1%	8.34	.76
5	.486	.1%	4.06	.76
6	.468	.1%	8.46	.78

*Regression equations are listed in the same order that they are presented in this section.

that all multiple correlations are higher than the simple correlations. Also, the residual variances of the regression equations are lower than the residual variances of the simple correlations, indicating that the multiple correlations are better predictors of vocational agriculture teacher success than are the sample correlations.

All multiple correlations are statistically significant, indicating that the relationships found cannot be attributed to chance alone. Regression equation number 3 appears to be the best predictor of vocational agriculture teacher success investigated in this study. In the third regression equation, the farm management test score, the dairy and beef cattle manipulative farm skills score, and the professional education test score were employed to predict the success of a vocational agriculture teacher, the criterion of success being the composite of individual rating scores.

SECTION IV

SUMMARY AND CONCLUSIONS

Statement of the Problem

THE PURPOSE of this investigation was to study the relationship between the success attained by vocational agriculture teachers and certain competencies that these vocational agriculture teachers possess.

Design of the Study

The subjects of this investigation were a random sample of sixty vocational agriculture teachers who teach in one teacher vocational agriculture departments of public high schools located in approximately the southern two-thirds of the state of Wisconsin.

The measure of teacher success was determined by ratings given to each vocational agriculture teacher by: (1) state supervisors of vocational agriculture, (2) high school principals, (3) high school pupils taught by the teacher, and (4) farmers in the local school community. A composite score of individual raters and a composite score of combinations of raters were used in analyzing the data.

The competencies of vocational agriculture teachers that were measured in this investigation were in the areas of technical agriculture, manipulative farm skills, and professional education. Technical agriculture competency was measured by administering tests of factual information in seven technical agricultural fields. These tests were in the areas of agronomy, dairy cattle, farm management, horticulture, poultry, soils, and swine.

Competency in manipulative farm skills was measured by a self-rating survey of manipulative farm skills that was completed by each subject. This survey of skills was limited to fifteen farm enterprises or areas as follows: dairy and beef cattle, poultry, swine, corn, meadows and pastures, small grain, gardening, potatoes, fruit growing, concrete, electrification, farm machinery and power, erosion control, home beautification, and farm plumbing.

The measure of professional education competency of vocational agriculture teachers was accomplished by administering the professional information section of the 1951 edition of the National Teacher Examination to each teacher participating in this investigation.

Using the two composites of scores for the criteria and the raw scores from the measures of competence, Pearson product-moment correlations and regression equations were calculated to determine the relationship between vocational agriculture teacher success and certain competencies possessed by these teachers.

Summary of Findings

All tests in technical agriculture were designed and developed in a similar manner. They all dealt with basic and fundamental factual material in their respective areas. Because of these similarities, the various tests of technical agriculture were assumed to have approximately equal difficulty and equal basic understanding requirements.

Percentage scores made on the tests in technical agriculture indicate that vocational agriculture teachers in Wisconsin, from which this sample was chosen, have a better knowledge of basic facts in animal enterprises than they have in other farming enterprises investigated in this study. These teachers appear to have less knowledge of basic facts in soils and farm management than they have in other areas of farm endeavor investigated in this study. Their knowledge of basic facts in agronomy and horticulture seems not as great as their knowledge of basic facts in animal enterprises. However, their knowledge of basic facts in agronomy and horticulture seems greater than their knowledge of basic facts in soils and farm management.

The mean score made by the vocational agriculture teachers on the professional education test was about thirteen points lower than the mean score made by persons who expressed a preference for teaching vocational subjects who took the nation-wide administration of the same test in 1951. This comparison must be made with caution since the number of vocational teachers used in this study is small compared with the number of persons expressing vocational teaching preferences who took the national administration of the test in 1951. Also, persons taking the national administration of the test in 1951 may have made special study and reviews in preparation for taking the test because of its use in teacher evaluation and promotion in some situations. No such special preparation was made by teachers participating in this study.

There was no particular relationship among the mean, high, and low percentage scores made on the various enterprises of the manipulative farm skills survey. This indicates a rather wide diversity of individual competencies in performing agricultural skills.

There were consistently low correlations between the manipulative farm skills scores and the technical agricultural test scores. These consistently low correlations indicate that there is very little relationship between a vocational agriculture teacher's knowledge of factual information in agriculture and his own judgment of his ability to perform manipulative farm skills.

All correlations between the professional education test scores and scores made on various enterprises of the technical agriculture test were positive, and a majority of them were statistically significant. There was a statistically significant correlation between the professional education test score and the technical agriculture test composite score. These correlations indicate that there is a significant relationship between a vocational agriculture teacher's information on professional education and his knowledge of basic facts in technical agriculture enterprises.

All correlations between professional education test scores and manipulative farm skill scores were negative except one, and none were statistically significant except one, which indicates that there is very little relationship between a vocational agriculture teacher's information on professional education and his own judgment of his ability to perform manipulative farm skills.

In order to facilitate the achievement of the purpose of this study, four null hypotheses were formulated. These hypotheses served as a guide in making basic judgments that were necessary in this study. The four null hypotheses tested in this study were:

1. There is no statistically significant relationship, as shown by coefficients of correlation, between a vocational agriculture teacher's knowledge of technical agricultural information and his success as a teacher.

2. There is no statistically significant relationship, as shown by coefficients of correlation, between a vocational agriculture teacher's proficiency in performing agricultural manipulative skills and his success as a teacher.

3. There is no statistically significant relationship, as shown by coefficients of correlation, between a vocational agriculture teacher's knowledge of professional educational information and his success as a teacher.

4. There is no statistically significant relationship, as shown by regression equations, between a vocational agriculture teacher's knowledge of technical agricultural information, proficiency in performing agricultural manipulative skills, and knowledge of professional educational information and his success as a teacher.

The first null hypothesis postulated above was rejected. It was rejected on the basis of statistically significant correlations between the composite of technical agriculture test scores and the composite of individual rating scores, and between the composite of technical agriculture test scores and the composite of combinations of rating scores.

The second null hypothesis postulated in this study was accepted. It was accepted because there was no statistically significant correlation between the composite of manipulative farm skill survey scores and the composite of individual rating scores, and between the composites of manipulative farm skill survey scores and the composite of combinations of rating scores.

The third null hypothesis postulated in this study was neither accepted nor rejected. It was left open for question and further study. This decision was made on the basis of a statistically insignificant correlation being found between the professional education test score and the composite of individual rating scores, while a statistically significant correlation existed between the professional education test score and the composite of combinations of rating scores.

The fourth null hypothesis postulated in this study was rejected. It was rejected on the basis of statistically significant correlations between the combination of competency scores and the composite of individual rating scores, and between the combination of competency scores and the composite of combinations of rating scores.

No coefficients of correlation and no regression equations investigated in this study accounted for enough of the total variance of scores to make them entirely satisfactory for predicting individual scores.

The best predictor of vocational agriculture teacher success found in this study was using the combination farm management test score, dairy and beef cattle manipulative farm skills score, and the professional education test score to predict the composite of individual rating scores. The technical agriculture tests were better predictors of vocational agriculture teacher success than were the manipulative farm skills, which had very little relationship to success, or the professional education test, whose exact relationship to vocational agriculture teacher success was not determined. The best single predictor of vocational agriculture teacher success was the farm management test.

Conclusions

The relatively low percentage scores made on the farm management and soils technical agriculture tests, the professional education test, and several of the manipulative farm skill enterprises leads one to question the adequacy of the curriculum of training of vocational agriculture teachers in these areas.

A vocational agriculture teacher needs a comprehensive knowledge of farm management and soils if he is to give maximum service to farmers. The farm management test correlates higher with success rating composites than any other single competency measure. This indicates that there is a closer relationship between a vocational agriculture teacher's success, as measured by composites of ratings, and his knowledge of basic information in farm management than there is between his success, as measured by the same criteria, and knowledge of basic information in any other technical agricultural enterprise investigated in this study.

The correlations between both soils test and the professional education test with success as measured by composite of ratings is statistically significant.

Findings in this investigation tend to substantiate findings by Rhoad (study referred to in Section II), concerning the ability of vocational agriculture teachers to perform manipulative farm skills. It cannot be assumed that vocational agriculture teachers adequately learn to perform manipulative farm skills on the home farms. All teachers participating in this study were farm reared except one, and yet, they do not judge themselves as being very proficient in performing manipulative farm skills. In view of this finding, it would appear wise for training institutions to devote more time to the training of vocational agriculture teachers and prospective teachers to perform manipulative farm skills more efficiently.

Implications for Further Study

As a result of findings in this investigation, the author wishes to make five recommendations for further study along the line of predicting vocational agriculture teacher success.

The first recommendation is that the farm management test, the poultry test, the soils test, and the swine test that were used in this investigation be given further study with a view toward making them, if possible, better individual and collective predictors of vocational agriculture teacher success. These tests of technical agriculture showed statistically significant relationships with vocational agriculture teacher success as measured by composites of raters. Through an item analysis one could determine which questions in these tests best differentiate the most successful teachers from the less successful teachers. In this manner one might develop an instrument that would better predict vocational agriculture teacher success.

The second recommendation is that the professional educational test used in this investigation be given further study with a view toward making it a valuable predictor of vocational agriculture teacher success. The value of the professional education test as a predictor of vocational agriculture teacher success was inconclusive in this investigation. As previously stated, the subjects used in this investigation made relatively low scores on the professional education test. It might be that vocational agriculture teachers are so laden with keeping conversant with technical information that they are forced to give professional information minimum attention. Then too they may depend more on practical, seemingly logical solutions than upon knowing established principles. An item analysis of the professional education test should indicate the type of professional questions that agriculture teachers answer correctly and which are the most discriminating questions. Such knowledge might aid in the development of an instrument in professional education that would be particularly suited for predicting vocational agriculture teacher success.

The third recommendation is that additional objective data such as age, teaching experience, college grade point average, credits in technical agriculture received in college, credits in professional education courses received in college, credits in general education courses received in college, and practice teaching credits be employed in a study such as this to predict vocational agriculture teacher success. If certain objective data suggested here were used in conjunction with the most predictive variables of vocational agriculture teacher success used in this study, improved prediction of vocational agriculture teacher success might be possible.

The fourth recommendation is that certain subjective data in the areas of human relationships and personality be employed in a statistical way to determine their importance as predictors of vocational agriculture teacher success. The somewhat unique responsibilities and duties that a vocational agriculture teacher has in a school and in his community might make his personal qualities very important to his success as a teacher. The use of such subjective data in conjunction with certain objective data might prove to be of great value for predicting vocational agriculture teacher success.

Finally, the author recommends further study of subjective ratings, and the possible use of residual pupil gains as success criteria in studies of vocational agriculture teacher success. The criterion by which teaching success is evaluated is of great importance. An objective criterion such as residual pupil gains might add much to studies of this kind. The fact that only one school subject, vocational agriculture, is being dealt with, and that the teachers being investigated have the same curriculum for use as a guide in teaching would facilitate the use of residual pupil gains as a criterion of vocational agriculture teacher success. It appears that composites of ratings give better results than ratings used singly. However, more study is needed in defining aspects of teaching to be evaluated and in objectifying subjective rating as much as possible. It is believed that the use of these recommendations will lead to better studies of the prediction of vocational agriculture teacher success.

BIBLIOGRAPHY

1. Administration of Vocational Education. Vocational Education Bulletin No. 1, General Series No. 1, Revised 1948 (Washington, D.C.: U. S. Government Printing Office, 1949).
2. Anderson, Clarence Scott. An Analysis of Certain Pre-Employment Records and Activities of Candidates for Teaching Vocational Agriculture, Ph.D. Dissertation, Cornell University, 1933.
3. Anderson, Harold M. A Study of Certain Criteria of Teaching Effectiveness, Ph.D. Dissertation, University of Wisconsin, 1952.
4. Armstrong, Watson. Relation Between College Grades and Success of Teachers of Agriculture in Kentucky, unpublished manuscript, University of Kentucky, 1933.
5. Barr, A. S. "Recruitment for Teacher Training and Prediction of Teaching Success," Review of Educational Research, X (June, 1940), pp. 185-190.
6. ——. "The Measurement and Prediction of Teaching Efficiency," Review of Educational Research, XIII (June 1943), pp. 218-223.
7. ——. "The Measurement and Prediction of Teaching Efficiency," Review of Educational Research, XVI (June 1946), pp. 203-208.
8. ——. "Measurement and Prediction of Teaching Efficiency," Review of Educational Research, XIX (June 1949), pp. 185-190.
9. ——. "The Measurement of Teacher Characteristics and Prediction of Teaching Efficiency," Review of Educational Research, XXII (June 1952), pp. 169-174.
10. ——. "The Measurement and Prediction of Teaching Efficiency: A Summary of Investigations," Journal of Experimental Education, XVI (June 1948), pp. 203-283.
11. Brown, James Frank. The Relationship of Personality Traits and Vocational Interest to Success in Teaching Vocational Agriculture, M.S. Thesis, Virginia Polytechnic Institute, 1940.
12. Carlson, Carol R., and others. An Educational Statistics Primer (Madison, Wis.: Dembar Publications, Inc., 1950).
13. Cook, Glen Charles. A Handbook on Teaching Vocational Agriculture (Danville, Ill.: Interstate Printing Co., 1947).
14. Coombs, Joseph Gleen. An Analysis of the Relationship Between Directed Experiences Obtained in Training and First-Year Teaching Performance for Teachers of Vocational Agriculture, and an Examination of the Procedure Used, Ph.D. Dissertation, Cornell University, 1951.
15. Cormier, Ulysse. The Evaluation of Some of the Influences Which May Relate to Success in Teaching Vocational Agriculture, M.S. Thesis, Virginia Polytechnic Institute, 1940.
16. Crissy, William J. E. "National Teacher Examinations," Phi Delta Kappan, XXIV (May 1952), pp. 353-356.
17. Domas, Simeon J., and Tiedeman, David V. "Teaching Competence: An Annotated Bibliography," Journal of Experimental Education, XIX (December 1950), pp. 101-218.
18. Efferson, Carlos Arthur. Pre-Training Records and Activities of Teachers of Vocational Agriculture in Louisiana as Related to Teaching Success, M.A. Thesis, Louisiana State University, 1940.
19. Fisher, Ronald A., and Yates, Frank. Statistical Tables for Biological, Agricultural and Medical Research, (New York: Hafner Publishing Co., 1948).
20. Flanagan, John C. "An Analysis of the Results from the First Annual Edition of the National Teacher Examinations," Journal of Experimental Education, IX (March 1941), pp. 237-250.

21. Floyd, John Claiborn. Pre-Training Records and Activities of Successful Teachers of Vocational Agriculture, Ph.D. Dissertation, University of Missouri, 1939.

22. Good, Carter V., Barr, A. S., and Scates, Douglas E. The Methodology of Educational Research (New York: D. Appleton-Century Co., 1941).

23. Hardway, Wendell G. Participating Experiences Obtained in Teaching and Subsequent Performance in Teaching for Teachers of Vocational Agriculture, Master's Problem, West Virginia University, 1952.

24. Johnson, Palmer O. Statistical Methods in Research (New York: Prentice-Hall, Inc., 1949).

25. Jones, Ronald DeVall. "The Prediction of Teaching Efficiency from Objective Measures," Journal of Experimental Education, XV (September 1946), pp. 85-100.

26. Lins, Leo Joseph. "The Prediction of Teaching Efficiency," Journal of Experimental Education, XV (September 1946), pp. 2-60.

27. Love, Harry M. A Study to Determine Some of the Qualifications Necessary to Success in Teaching Vocational Agriculture, M.S. Thesis, Virginia Polytechnic Institute, 1932.

28. McClay, David Ross. A Determination of the Relationship Between Experience Obtained in Training and Subsequent Performance in Teaching for Teachers of Vocational Agriculture and an Examination of the Process Used, Ph.D. Dissertation, Cornell University, 1951.

29. Montross, Harold Wesley. Temperament and Teaching Success, Ph.D. Dissertation, University of Wisconsin, 1952.

30. Peters, Charles C., and VanVoorhis, Walter R. Statistical Procedures and Their Mathematical Bases (New York: McGraw-Hill Book Co., 1940).

31. Rhoad, Claude Elton. A Study of the Relation of Selected Pre-Service Achievement and Experience Factors to Success in Teaching Vocational Agriculture in Ohio, M.A. Thesis, Ohio State University, 1937.

32. ———. A Study of the Comprehensiveness of Abilities in Technical Agriculture Attained by Prospective Teachers of Vocational Agriculture in Ohio Previous to Their Entrance into Student Teaching, Ph.D. Dissertation, Ohio State University, 1943.

33. Ryans, David G. "The Use of the National Teachers Examinations in School Systems," Educational Administration and Supervision, XXXV (February 1949), pp. 65-88.

34. Shaw, Byron T. "Impact of Research in Agriculture," Phi Delta Kappan, XXXV (October 1953), pp. 11-15.

35. Summaries of Studies in Agricultural Education. Vocational Education Bulletin, No. 180 (Washington, D.C.: U. S. Government Printing Office, 1935).

36. Summaries of Studies in Agricultural Education. Supplement No. 1 to Vocational Education Bulletin No. 180 (Danville: Interstate Printers and Publishers, 1943).

37. Summaries of Studies in Agricultural Education. Supplement No. 2, Vocational Division Bulletin No. 237 (Washington, D.C.: U. S. Government Printing Office, 1948).

38. Summaries of Studies in Agricultural Education. Supplement No. 3, Vocational Division Bulletin No. 242 (Washington, D. C.: U. S. Government Printing Office, 1950).

39. Summaries of Studies in Agricultural Education. Supplement No. 4, Vocational Division Bulletin No. 246 (Washington, D.C.: U. S. Government Printing Office, 1951).

40. Summaries of Studies in Agricultural Education. Supplement No. 5, Vocational Division Bulletin No. 248 (Washington, D. C.: U. S. Government Printing Office, 1952).

41. Summaries of Studies in Agricultural Education. Supplement No. 6, Vocational Division Bulletin No. 251 (Washington, D. C.: U. S. Government Printing Office, 1953).

42. Sutherland, S. S. "Can We Predict Teaching Success?" Agricultural Education Magazine (August 1937), pp. 35, 38.

43. National Teacher Examination, Bulletin of Information (Princeton: Educational Testing Service, 1953).

44. Thomas, Godfrey H. The Factorial Analysis of Human Ability (New York: Houghton Mifflin Co., 1939).

45. Torrence, Andrew P. How the Vocational Agriculture Curriculum of Wisconsin Contributes to the Ten Imperative Needs of Youth as Listed by the Educational Policies Commission, unpublished manuscript, University of Wisconsin, 1951.

46. Vocational Education in the Years Ahead. United States Office of Education, Vocational Division Bulletin No. 234 (Washington, D.C.: U. S. Government Printing Office, 1945).

47. Walker, Helen M. Elementary Statistical Methods (New York: Henry Holt and Co., 1943).

THE ROLE OF EVALUATION IN THE IMPROVEMENT OF INSTRUCTION IN INSTITUTIONS OF HIGHER LEARNING

JAMES E. GREENE, Sr.
University of Georgia

IN APPROACHING the present task of discussing the role of evaluation in the improvement of instruction, an attempt will be made to provide brief, tentative, and perhaps somewhat personalized answers to the following questions:

1. What is the essential rationale of a philosophy of evaluation of instruction?
2. What evidences can be cited to indicate that emphasis upon improved evaluations of instruction tend to improve the general quality of instruction?
3. How can an institution of higher learning improve its program of evaluation of instruction?

For our present purposes, the term evaluation will be used as referring to (1) a process, and (2) a product. Considered as a process, evaluation refers to the procedures by means of which value judgments are made concerning a condition or procedure having implications for effective instruction. Considered as a product, evaluation refers to the nature and characteristics of the value judgments themselves.

The Rationale of a Philosophy of the Evaluation of Instruction

If emphasis is to be placed on the rationale of evaluation in the improvement of instruction, the first logical step involved in the development of our theme would consist of a logical analysis of the basic constituent elements involved in effective instruction. For our present purposes it is held that instruction is likely to be effective in proportion as:

1. Each objective of instruction is clearly understood and fully accepted by the learner and is uniquely appropriate to his abilities and needs.
2. Each of the learning situations comprising the instructional program (i. e., teacher and learner activities, methods and materials of instruction, climate of learning, etc.,) is nicely adapted to the efficient attainment of one or more of the objectives sought.
3. Teacher and learner each secure, analyze and interpret evidences concerning the total efficiency of the instructional process.

The overall task of evaluation in relation to the improvement of instruction would appear thus to consist of finding sound and defensible answers to such questions as the following:

1. To what extent has each learner made or failed to make satisfactory progress toward the attainment of each of the objectives of instruction originally formulated and accepted by both teacher and learner?
2. To what extent were the original objectives of instruction appropriate or inappropriate to each learner?
3. What revisions, if any, of the objectives of instruction would appear to be desirable?
4. To what extent were the instructional procedures (i. e., materials, methods, etc.,) appropriate to the attainment of the specific objectives sought?
5. What revisions, if any, of the instructional procedures appear to be desirable?

If evaluation is to serve the purposes thus far stated or implied, it is requisite that evaluations conform to certain basic principles and that certain basic procedures in evaluation are rigorously observed.

What basic principles are involved in the evaluation of instruction?

Conformity to the following principles is held to be basic to the development of a sound program of evaluation:

Evaluations of instruction must have a defin-

---

*This article and the succeeding three articles (including an article each by Drs. Ralph Berdie and William Coleman, which have been published elsewhere) were presented before the symposium jointly sponsored by the AERA and Section Q of the AAAS in Atlanta, Georgia, December 27, 1955.

ite frame of reference. That is, evaluations are to be made in terms of the specific values, goals or objectives of teacher and learner if such evaluations are to be meaningful and helpful. Evaluative procedures which fail to take into account the self-determined goals of the individual being evaluated are not only of limited or doubtful worth; they may actually be pernicious in their distorting or coercive effects.

Evaluations should be continuous and cumulative. Any evaluation as of a given moment is likely to be erroneous or incomplete in certain particulars; and, in any case, evaluations as of a given moment cannot show changes in status. Furthermore, since the goals of individuals change as conditions change, it is necessary for sound evaluative procedures to take into account these changes in individual goals for the reason that the prognostic significance of a given evidence of excellence may vary from time to time in such a way as to make that evidence of greatly unequal value on different occasions.

Evaluations should be comprehensive and conclusive. Since in any given situation multiple goals of learning usually are being sought it is necessary to secure evidences of the extent to which each of the goals of instruction has been or is being attained. It is clear that in certain instances given goals may be mutually exclusive, or that progress toward a given goal may be offset by losses with respect to other goals. While individuals should be evaluated on the basis of a "total pattern" of characteristics of desired excellence, it should be recognized that superiority in some characteristic may be regarded as compensating, to some extent, for deficiencies in other respects.

The primary purpose of all evaluative procedures is that of encouraging and promoting improvement with respect to the condition or activity being evaluated. Consequently, all evaluative procedures should permit of and provide for specific, constructive suggestions for improvement. In general, effective evaluation will make slight and incidental, if any, use of inter-individual comparisons; rather it will compare the record of the given individual at successive time intervals and always in the light of his own unique needs and purposes. Evaluative procedures should generally avoid the use of the concept of "fixed standards" of learner attainment—since in the past this concept has often implied "an identity of individual learner needs and purposes." Sound evaluative procedures will emphasize flexibility and optimum growth in terms of differential purposes.

In view of the essential nature of evaluation (i.e., making value judgments of the progress being made toward specific objectives) and of the basic purpose of evaluation (i.e., improvement of a condition or a process) it is clear that sound evaluative procedures will involve the active participation of the persons primarily concerned in the evaluation. That is, effective evaluation will emphasize self-appraisal activities. However, this statement does not exclude the use of outside persons of specialized skills in evaluation who, on invitation, may be prepared to serve in a resource or consultant capacity on various technical aspects of the evaluation program. Such consultants would likely be of most value in:

1. suggesting clarifications or sub-analyses of objectives such that it would be possible to secure appropriate evidences relative to each sub-objective;

2. suggesting possible techniques and procedures for securing appropriate evidences of goal attainment;

3. suggesting cautions to be observed in relating the obtained evidences to the objectives sought.

What basic procedures are involved in the evaluation of instruction?

Evaluation of instruction involves the following fundamental processes:

Adequate Analysis of Objectives—That is, objectives must be analyzed in such terms that it is possible to secure appropriate evidences of the extent to which such objectives are being realized. It is fruitless to formulate an objective, however worthy, unless it is possible to analyze it in such a way that evidences can be obtained to show the extent to which it is being realized. If the objectives of instruction are stated in vague or general terms it may be difficult to secure evidences showing the extent to which they have been attained.

Securing Appropriate Evidences of a Status Product, or Process—All evidence items must have a clear and definite (even if indirect) relationship to the objectives. Evidences may involve either subjective or objective data, or both. Commonly used forms of evidence include: scores on objective or essay tests, products created by the learner, rating scales, interviews, anecdotes, sociometric devices, observational data, etc. It is clear that the type of evidence required in a given situation will be intimately related to the specific objective, i.e., the nature of the objective will determine the type of evidence required by the situation. Obviously, the evaluation will be sound only in proportion to the quantity and quality of the evidence upon which it is based.

Relating the Obtained Evidences to the Objectives Sought—This step involves critical, reflective thinking in answering the query, "To what extent do the available evidences indicate that the desired objectives have been or are being effectively realized?" It is again obvious that the evaluations obtained will be dependent upon the soundness of the analytic procedures employed in arriving at them.

Empirical Justification of an Emphasis on Improved Evaluations of Instruction

The empirical data to be here cited will be limited to a single institution. As based on their experiences as former members of the Educational Advisory Staff of Air University, Maxwell Air Force Base, Alabama, in a series of previous papers* the writer and certain of his professional associates have attempted to demonstrate that systematic emphasis on the improvement of evaluative procedures not only resulted in improved evaluations per se but also contributed to improved objectives of instruction and to improved instructional procedures.

One of these previous papers dealt with the sub-hypothesis that the quality of instruction could be improved by means of (1) improving objective test items, (2) improving the measurement of intangibles, and (3) utilizing student participation in the evaluation of instruction.

The evidence that emphasis upon the improvement of objective test items actually resulted in improved evaluations consisted of considerable data showing a measurable and statistically significant improvement in such test items as measured by increased validity coefficients and by an improved functional content of test items on successive tests in the same area of instruction. It was further found that improved validity coefficients and improved functional content of test items were accompanied by increased "face validity" of tests, increased student satisfaction with the "fairness" of tests and an increased critical concern by instructors and students relative to the appropriateness of the instructional objectives and the teaching procedures employed.

Time limitations preclude the presentation here of detailed evidence indicating that emphasis on the evaluation of intangible outcomes and on the participation of learners in the evaluative process was accompanied by improved evaluations per se as well as improved objectives and improved instructional procedures. It may be summarily stated, however, that these evidences included such concrete data as changes in the official statements of the objectives of particular units of instruction, changes in the amounts of time devoted to particular units of instruction, and changes in the amount of time devoted to particular instructional procedures (i. e., lecture, discussion, demonstration, etc.). The evidence also included the systematically analyzed judgments of members of the Educational Advisory Staff, of the instructional staff, and of the students themselves.

Institutional Improvement in the Evaluation of Instruction

Three general suggestions are made relative to improved evaluations of instruction in individual institutions of higher learning:

1. Recruitment—In the employment of instructional personnel special emphasis should be given to the competencies in evaluation of prospective employees.

2. In-service training—Considerable evidence indicates that an in-service training program can bring about notable improvements in the evaluation competencies of instructors.

3. Specialized evaluation services—Experience in the armed services, in many secondary schools and in a few institutions of higher learning appear to warrant the belief that provision by the Central Administration of specialized evaluation services, available on a voluntary basis to members of the instructional staff, tends to bring about not only marked improvement in the evaluation of instruction per se but also facilitates improved instructional objectives and improved procedures of instruction.

---

*Other published papers in this series include the following: James E. Greene, "The Evaluation of Instruction in Air University," Journal of Psychology, XXV (1948), pp. 279-297; James E. Greene and Warren G. Findley, "Evaluative Procedures for the Improvement of Instruction," Educational Record, XXX (January 1949), pp. 33-44; Hugh F. Seabury and Kenneth R. Williams, "Air University Trains Instructors," Higher Education, IV (September 15, 1947), pp. 15-18; Kenneth R. Williams and Alder M. Jenkins, "Improving Instruction in Institutions of Higher Education," Educational Record, XXIX (April 1948), pp. 145-161.

THE ROLE OF EVALUATION IN THE IMPROVEMENT OF PLACEMENT AND FOLLOW-UP IN INSTITUTIONS OF HIGHER LEARNING

PAUL I. CLIFFORD
Atlanta University
Atlanta, Georgia

THE PURPOSE of this paper is to examine the role that evaluation can play in the improvement of the placement and follow-up aspects of personnel services in institutions of higher learning.

Conceptual Framework

In order to adequately identify the role of evaluation in these areas, it is first necessary to provide a frame of reference or a set of terms into which the discussion can be cast. This conceptual framework includes precise definitions concerning these terms and their interrelations which have been accepted by the writer. Evaluation, in the specific sense used in the present paper, refers to the various technics and processes which can be utilized to determine the degree to which definitely stated and clearly defined objectives of the placement and follow-up services are being attained. Thus, evaluation includes any means by which valid evidence may be collected and interpreted for the purpose of judging the efficacy or value of present procedures that are being employed to attain the specific ends of placement and follow-up.

Although placement and follow-up are sometimes considered as distinct services, they are inseparable in both theory and practice. Placement may be defined as a multi-phasic process which aids an individual to make those decisions concerning an educational or vocational next step which will increase the probability that he will progress satisfactorily from one educational or vocational experience to another. Follow-up is the evaluative phase of placement and all the other antecedent services involved in the guidance and instructional programs of a given institution.

Placement, in its educational phase, has for its major objective the matching of an individual student and an array of particular educational experiences in such a way that the student will likely attain maximum benefit. In its vocational phase, the aim of placement is to assist the individual to find his appropriate place in the world of work. Appropriate, as used here, means a job, a vocation, or a profession which (1) is consonant with one's interests and other emotionally toned preferences, (2) on a level with the configuration of one's general and specific mental abilities, (3) makes use of one's achievements and past acquisitions of skills, (4) makes no physical demands upon the individual that he cannot satisfy, (5) is personally satisfying or rewarding to the individual to the degree that it continually contributes to higher levels of personality integrity and stability, and (6) makes a significant contribution to the improvement of the social order with which the individual is in interaction. It must always be borne in mind that placement is individually oriented. Its principal focus is always the individual who is being placed. Placement should not be confused with either recruitment or selection. Recruitment, in contrast to placement, is concerned primarily with meeting the needs of employers for workers or of educational institutions for students, while selection is the process, by which, from among many applicants, certain ones are picked for given and limited educational or vocational experiences. It should be apparent that since the foci of the placement, selection, and recruitment services are different, the three terms should never be used synonymously or become confused in our thinking.

Follow-up is conceived of as the continuous process of determining, by any valid means, the degree to which an individual's educational or vocational choices are rewarding or satisfying to him in terms of the six criteria of appropriateness identified above. Follow-up services should not be thought of as applicable only to the graduates of an institution of higher learning but should be conceptualized to include the appraisal of the degree of educational or vocational adjustment of those who have withdrawn before graduation. It should be axiomatic that both placement and follow-up services should be employed with former students as long as there is need on the part of the individual for these services.

Basic Assumptions

At this point, it is necessary to present the basic assumptions which the writer has accepted in the preparation of this paper. The first assumption is that our institutions of higher learning are genuinely concerned with developing to the fullest extent possible the potentialities possessed by each student which they admit. If this assumption is valid, it follows, then, that our educational enterprises have the moral responsibility of using the totality of their resources not only to help their former students to take the next appropriate educational or vocational step, but also to determine as validly as possible whether the step taken was the one most likely to eventuate in satisfaction to the individual and benefit to society.

The second basic assumption is that the only means by which a given institution of higher learning can determine the degree to which its placement and follow-up services are attaining their objectives is to carry on planned, valid, comprehensive, and continuous appraisals of these programs.

The third basic assumption is that in order to evaluate placement and follow-up, the objectives of these services must not only be clearly stated in terms of student behavior, but the particular areas of functioning in which the behavior is expected to take place must also be identified. This assumption required us to think of the objectives of both placement and follow-up as acquisitions of behavior or learnings on the part of the individual student which will materially increase the probability that he will make and continue to make satisfying and rewarding educational and vocational choices after leaving the educational institution he is presently attending.

The fourth basic assumption is that the data which are derived from evaluative processes which have as their locus the behavior of the individual student, must be referred, in the final analysis, to the specific opportunities or tasks to be performed in educational institutions and in the world of work. This reference, or the determination of the appropriateness of students on one hand and vocational and educational opportunities on the other, requires evaluation at both poles of the individual—vocational or educational spectrum.

Roles of Evaluation and Follow-Up

In view, then, of the conceptual framework and the basic assumptions of this paper, it is possible to identify the roles of evaluation in placement and follow-up in institutions of higher learning as at least four in number.

The first of these is to provide basic information about the individual student in terms of his general and specific mental abilities, his school achievement, his interests, his personality integration, his socialization, his physcial status, and any other attributes which are relevant to success in further educational endeavors or in the world of work. In playing this role evaluation makes it possible for the psychologist or the personnel worker to synthesize a set of descriptions of the individual's behavior in these various domains in such a way as to get a holistic psychological picture of the individual's functioning as a totality.

The second role of evaluation in placement and follow-up is to provide valid and usable information about the approximately 30,000 occupations in the United States and about the many educational opportunities available to students who graduate or withdraw from our institutions of higher learning. The processes of evaluation must be expanded to include the continuous gathering of valid data about the world of work in terms of industrial pursuits and processes, job trends, labor supply and demand, and the expectations that employers have concerning the total functioning of their employees. Unless evaluation plays the second role just identified, its first role will be of little avail. Unless a personnel worker understands the nature of the occupational and educational world, his understanding of the individual which he is ostensibly helping to adjust to vocational or educational expectations is mere sophistry.

The third role of evaluation in placement and follow-up is to provide a basis for the confluence of valid information about the individual and accurate data about the vocational and educational worlds. Evaluation provides the only means by which these two kinds of information can be reduced to a set of terms which, at once, are utilitarian and capable of providing the stimulation for progressively deeper and broader syntheses of those seemingly polar considerations.

The fourth role of evaluation in placement and follow-up is to determine in as valid and comprehensive a manner as possible the degree to which students have acquired the kinds of behavior which are associated with educational and vocational adjustment. In order for this to be accomplished our institutions must make periodic, longitudinal studies of at least a representative sample of their students' further educational or vocational endeavors.

Because of individual differences and the variability of human beings on the one hand and the increasing heterogeneity and complexity within the educational and vocational worlds on the other, evaluation must be called upon increasingly to help determine the appropriateness of the one for the other.

It is quite apparent that the task which the writer has set for evaluation in improving placement and follow-up is a gigantic one.

We have been attempting, for over half a century, to improve our methodological approaches to more valid and comprehensive descriptions of the behavior and functioning of human beings. We have barely scratched the surface, however, in determining how we might use these data that we can collect in order to match men and opportunities. This, to the writer, is the challenge of the future.

The notion that evaluation can play a role in identifying the nature of vocational opportunities in America is a relatively novel one. It should be possible for psychologists, personnel workers, job analysts and others to pool their competencies in such a way as to produce a classification of occupational opportunities in terms of (1) the primary personal attributes which workers should possess if they are likely to be successful in one or more of the various occupational families, (2) the degree of general and special mental abilities which students must possess if they are to learn to function properly within a given family of occupations, and (3) the kinds of educational achievements which are necessary for the student to enter one or more of the various occupational families. It should be pointed out here that, at best, we have only made very rough beginnings at estimating or guessing what the many worker constellations of characteristics, traits, or attributes required for success in our vastly complicated world of work are. If we are to improve upon these best guesses, we must begin the tremendous task of applying all of our modern technics of measurement, evaluation and assessment to this problem which will continuously become more difficult.

Concluding Statement

In conclusion, the writer would like to point out that one consequence of the wide acceptance of the four roles that evaluation can play in placement and follow-up, would be the saving of great sums of money by business. It has been estimated that the misplacement of a single employee in a business enterprise can cost from $50.00 to over $20,000.00 depending upon the level at which the misplaced person is functioning. It is presently beyond our capabilities to translate into dollars and cents the loss which business enterprises suffer in the areas of general employee morale, supervisory and administrative difficulties, high employee turnover, and diminished customer goodwill because of poor placement policies. The price paid in human unhappiness and in vocational mal-adjustment is inestimable.

Another consequence of the acceptance of the multi-dimensional roles of evaluation in placement and follow-up as identified in this paper could be the stimulation of more rigorous and precise research within the general field of the behavioral sciences and the specific fields of psychology and education. Only as research suggests the answers as to how we can better understand man on the one hand, and the world in which he lives on the other, can we attempt to effect a more satisfying rapprochement between the two.

THE ROLE OF EVALUATION IN SPECIAL SCHOLARSHIP SELECTION PROGRAMS

WESLEY W. WALTON
Educational Testing Service
Princeton, New Jersey

TIME WAS when the award of a scholarship was a matter strictly between a youngster and a college. Things are no longer that simple. Scholarship selection is rapidly becoming a big business. Third parties, fourth, and fifth now are involved. High school counselors screen and test candidates. Principals and headmasters execute candidate rating scales and furnish secondary school records. Testing bureaus in colleges and elsewhere administer secure testing programs. Selection committees of college admissions deans dive into seas of data to reach considered judgments as to who should receive scholarship awards and who should not. Financial analysts determine how much is needed to keep the scholar's budget in the black; and corporate sponsors furnish the funds necessary to move sponsored scholars happily on their way to college and through it.

How does one filter out of a group of eager and notably able scholarship aspirants numbering 60,000, the 600 who can make best use of higher education? This was the problem faced in the National Merit Scholarship Program in 1955-56. Or, by what means may we best be assured that the group of 16,000, as eager and as notable as those mentioned above, but with their sights on General Motors National Scholarship awards, are rendered down most equitably to 100 scholarship recipients? The latter problem has been faced for two years in the General Motors National Scholarship Plan.

Let us, for a moment, develop an hypothesis that young people in America come to a sophisticated understanding of the competitive system at a tender age. Interest in competing for scholarship opportunities is high. Further, that the talented among them seem to adjust readily to their own version of such a system. They are willing to take their chances. Finally, that in standing up to a nationally competitive situation, the candidate's whole concern is directed to a commendable performance as a scholarship candidate. He is so engrossed in the task that it is only in passing that he may "estimate" the odds against his winning.

Let us now add to the picture factors which may better be assumed than hypothesized.

1. There are not enough scholarships to go around to those who need and deserve them.
2. A substantial, though small and inadequate, number of young people who need and deserve scholarship aid will come through a scholarship competition successfully. Their accomplishments will long redound to the benefit of their benefactors, and they will contribute their bits to society on a higher plane than might otherwise have been the case.
3. Many who need and deserve scholarship aid will come through a scholarship competition unsuccessfully, and may perforce be kept from going to college as a result.

These groups of considerations, as they relate to special scholarship selection programs, may bring into focus several "insomnia breeders" experienced by people working in this aspect of evaluation. How may a scholarship selection committee be sure that it is being fair and just in removing a given candidate from further consideration? How does one distinguish minute but critical differences among the very top of a given ability grouping? What elements go into the selection of the very best scholarship risks?

At least partial answers to these questions may be found in the 1955 selection proceedings of the General Motors National Scholarship Plan. Scholarship selections were made by a nationally representative committee of college educators and secondary school men. The high caliber of its membership may be illustrated by citing the combination of President and Admissions Dean from all of the following: Vanderbilt, Notre Dame, Bryn Mawr, MIT, Dartmouth, Stanford, University of Michigan. The Principal of Phillips Exeter Academy, and the Director of Instruction for the Denver City Schools completed the membership. Let me now outline the selection procedures they used and then share with you their answers to three questions over which we had been loosing sleep.

At the outset the CEEB Scholastic Aptitude Test yielding separate verbal and mathematical scores on a 200-800 scale was administered to 13,112 high school students in January and March 1955. After scoring the test papers, total verbal and mathematical scores were computed for all candidates and candidates were ranked on this basis. A cutting score of 1345 (675 verbal or mathematical) was set to yield at least 1000 semifinalists for the available 100 awards. Adjustments to retain the original ratio between the sexes and to assure adequate geographic cover-age brought the semi-finalist population to 1414.

The next step toward selection involved the use of a weighted composite made up of twice the verbal score plus the mathematical score (with a weight of 2), the candidate's high school class rank relative to class size (with a weight of 3), a scaled rating by the school principal on the candidate's academic ability (with a weight of 1), and a scaled rating by the school principal on the candidate's personal and social adjustment (with a weight of 2). Such a composite was computed for the 1414 semi-finalists, and the ranking of candidates within the state accordingly led to the choice of about 500 finalists. At this point, to assure fairness and equitable handling, two categories of those semi-finalists falling below the cutting composite score for finalists were reviewed. Those cases with very high SAT scores and those identified as class valedictorians were studied and included in the finalist group if it seemed propitious to do so. The finalists, with these additions, number 585.

The final step for the selection of the 100 winners was built on subjective, intensive reading by the selection committee of the 585 finalist case files. Each file contained two descriptive forms, one completed by candidates about themselves and the other on their behalf by principals. Committee members worked in teams of two, each reading folders independently, and, on assessing the candidates' personal qualities, made a joint evaluation on the following four-point scale: (1) select, willing to defend, (2) select, (3) borderline between winner and non-winner, and (4) non-winner. These four points were used as ratings and were given numerical values of an order comparable to the scale scores used to rank finalists. Thus, subjective "clinical" evaluation had equal weight with objective "actuarial" evaluation. The sum of these evaluations then became the basis for the final ranking of candidates. After resolving ties and accommodating unfilled state quotas, the Committee recommended the top 100 for awards and 152 as alternates. Of the 100 General Motors National Scholars for 1955, 82 were on the list of 100 recommended for award, 18 were drawn from the alternate list.

The foregoing experience has suggested several considerations which may be regarded as partial answers to the questions previously raised.

How may committees be sure of fairness in removing a candidate from further consideration as a scholarship competitor?

1. Use a scheme of successive hurdles with at least some hurdles having multiple criteria.
2. Pass over each hurdle a large number of remaining competitors. (Large in comparison with the number of available scholarships.)
3. Provide by-passes to feed back into the competition those cases for whom "composite predictions" may mean injustice.

How does a scholarship committee make distinctions among highly competent competitors for scholarship?

1. Eliminates as rapidly as possible those who with fairness may be eliminated so that energies may be directed on finite evaluation among those of extreme competence.
2. Uses test instruments which will provide a spread of scores at the top end of the distribution.
3. Ties criteria tightly to the purposes set by the donor group and orders the criteria in accordance with the relative importance to the group, and weights evaluation accordingly.
4. Brings a wide variety of evidence—test and non-test—to bear on focusing differential and distinguishing traits. Draws this evidence from the candidate through personal data sheets, from his principal and teachers through a secondary school record and staff rating sheet.
5. Uses a dual approach involving both objectively predictive and clinically analytical types of activity.

What goes into the selection of best scholarship risks?

1. Adequate means for reducing a varied population having a wide geographic spread, varied academic and personal characteristics to a basis for comparability.
2. Systematic analysis of known predictive factors as they relate to the scholarship candidate and his chances for college success. Reduction of the prediction to quantitative terms.
3. Intense study of secondary school level of performance, academic, extra-curricular, community and otherwise, and an estimate of what this means in terms of college and post-college activity.
4. Determination of an "index of likely success" against which to measure a number of college-going potentials and from which to break into the open those who are best risks for sponsored scholarships.

THE ROLE OF EVALUATION IN THE DEVELOPMENT AND IMPROVEMENT OF REGIONAL PROGRAMS OF HIGHER EDUCATION

A. J. BRUMBAUGH
Southern Regional Education Board
Atlanta, Georgia

EVALUATION PLAYS a very important part in the development and improvement of regional programs in higher education. In this paper I shall discuss the subject within the general framework of the operations of the Southern Regional Education Board. The activities of the Board constitute a systematic and coordinated approach to the development and improvement of higher education in the fourteen states in the South. The concept of evaluation as it applies to the activities of the Board involves less of precise measurement and testing than is indicated in the preceding papers. It involves the identification of higher educational needs in the region, the determination of existing facilities to provide for the needs and proposals for bridging the gap between the needs and present facilities.

To consider the role of evaluation in the setting of the Southern Regional Education Board it is necessary to first summarize the major functions of the Board. Briefly stated, they are: (1) to serve as a clearinghouse for information on regionally significant activities in higher education; (2) to provide facility and staff for assessing needs and developing programs to meet needs of higher education in the South; (3) to administer inter-state arrangements for regional education services in institutions; (4) to act as fiscal agent in carrying out such arrangements; (5) to serve as a research facility on institutional and regional problems related to the improvement of higher education; and (6) to assist in providing consultative services to states and institutions on problems of major regional significance related to the improvement of higher education in the South.

A paper of considerable length would be required to discuss adequately the role of evaluation in relation to each of these functions. In order that I may stay within the limits of the time allotted I shall focus my presentation primarily on the second and fifth of the functions I noted; namely, assessing needs and developing programs to meet needs of higher education in the South and to serve as a research facility on institutional and regional problems related to the improvement of higher education. I shall also note briefly the necessity for appraising and marshalling social forces in order that programs of action may be effected.

To assess the needs and develop programs to meet the needs of higher education in the South is an evaluative task of major magnitude. It includes research to project as accurately as possible population growth and economic development as a basis for forecasting undergraduate and graduate enrollments and for determining the need for technically and professionally trained personnel in the light of economic changes. From these facts are determined the needs for new faculty members, new educational programs, increased physical facilities, and additional financial resources. The approach to a problem of this kind is complicated by the variables involved. First of all the statistical techniques employed must be defensible or else all of the conclusions derived from the data will be open to question. Second, certain assumptions whose validity has not been demonstrated must be made. For example, it is estimated that the Southern region faces a deficit of between 27,000 and 28,000 college faculty members by 1965. This is an evaluative judgment based on statistical evidence. The evidence includes the projected growth in enrollments in colleges and universities of the South, the faculty-student ratio, the teaching loads typical of Southern institutions, and the potential output of college teachers by the universities of the South. There are other assumptions too that have to be made—such as the rate of student migration into and out of the South and the percentage college age youth who will enter college. The validity of some of the assumptions has not been established, therefore the conclusions must remain tentative.

The Board is also called upon to determine needs for programs in various professional fields. For example, within recent years the question arose in one state as to the need for a school of architecture and in another state as to the need for a veterinary medical school. In each instance the Board was asked to assess the needs on a regional basis and to advise the states concerning the proposed programs. In both cases it was shown that the regional facilities for training people in these fields were adequate and that the states could much more economically use regional facilities than provide duplicate programs of their own. During the first seven years of the Board's operation more than thirty professional, technical, and graduate fields have been explored. To assess needs in these fields the Board must endeavor to find answers to many questions that will throw light on the current status of the field under consideration and the future needs in the field. Such questions as the following are illustrative:

1. What is the potential demand for professionally trained personnel in this field?
2. What is the potential student demand in the field?
3. What is the present student capacity of institutions in the region offering programs in the field?
4. What is the quality of existing programs in the field?
5. To what extent do existing programs in the field provide complementary or duplicate opportunities for specialization?
6. What cooperative arrangements among institutions and states would allow complementary strength of institutions to be combined or developed?
7. What additions, if any, are needed to improve the quality or to increase the capacity in the field in order to serve the needs of the region?
8. How can the interests and resources of industry, foundations, state and federal agencies be correlated with the special field so as to improve the quality of research and teaching and broaden the range of institutional services?
9. What previous experience of institutions and government can be drawn upon in assessing current needs in the field?

In assessing needs the first step is to identify the field of specialization to be considered, then to formulate the kinds of questions just noted. Priorities must also be established among the fields under consideration. The priorities of subject matter fields arrived at by the Southern Regional Education Board are determined largely on the basis of which fields will most likely elicit a response from the colleges and universities, state governments, and the general public. Once the fields to be assessed have been determined, the next step is to decide on the procedure to be followed in answering the questions. This phase of the evaluation can best be presented by referring to specific instances in which the process has been completed. One of the best illustrations is in the field of nursing education.

This project grew out of an action taken by the Southern Regional Conference of State Leagues of Nursing Education at its meeting in Atlanta on November 11, 1950. The motion passed at that meeting read:

> That the Board of Control for Southern Regional Education be asked to appoint a commission on nursing education as soon as possible for the purpose of studying nursing education and identifying the ways, if any, in which regional arrangements would be beneficial.

Other nursing groups gave strong support to this request. In response to the request the Southern Regional Education Board appointed a Committee on Nursing Education. The functions of this committee were as follows:

1. To assess regional needs for expansion and improvement of nursing education through
    a. Identifying needs for nurses able to execute various responsibilities;
    b. Identifying kinds and capacities of training facilities needed to meet these needs;
    c. Identifying needed research and other activities;
    d. Defining the gaps between present facilities and needed facilities, both in scope and quality.

2. To propose sound and feasible means of meeting identified regional needs through
    a. Action by the nursing and medical profession;
    b. Action by individual states or institutions;
    c. Action by groups of states or institutions.

The committee appointed by the Board had a two-day meeting in October 1951, at which were discussed the regional needs for nursing education, including the need for research which is recognized to be of fundamental importance to the sound development of the profession. From this meeting emerged several suggestions:

1. That regional arrangements for nursing education providing for graduate study to the level of the master's degree be considered by representatives of collegiate schools of nursing meeting with representatives of the Board.
2. That the same group also consider the possibility of establishing a center in the Southern region which would conduct research in nursing, act as a clearinghouse for research, and assist in preparing persons to do research in this professional field.

At the same meeting it was recommended that a competent person be employed to study the needs more thoroughly and to determine, after visits and the compilation of data, which universities seem at present best qualified to serve the region by offering graduate education in nursing, and also which universities seem best prepared to direct research, service, and education. The professional staff person appointed in accordance with the Committee's recommendation, in consultation with representatives of various professional groups, prepared a questionnaire designed to discover which of the universities in the region were interested in offering such services as were contemplated and what characteristics of these schools appeared to support their readiness to begin such service. The data presented by the universities which indicated such an interest were analyzed, summarized, and studied. Subsequently, the professional staff member on the project visited each school for the purpose of testing by direct observation and interviews the findings derived from the initial questionnaire study. The findings and conclusions thus reached were presented in a conference of representatives of the institutions having basic degree programs in nursing in 1952. From a list of eleven schools that initially expressed an interest in providing graduate programs in nursing or in establishing a research institute, six were finally selected. The criteria employed in making the final selection of these institutions were:

1. An institution of higher education of recognized academic and professional standing;
2. Institutional interest in graduate work as evidenced by an existent graduate council, dean and faculty and provision for inter-departmental communication.
3. Availability of a reference library of ample size and quality according to accepted standards, including its building.
4. Evident appreciation of the educational and health needs of the Southern region and a feeling of institutional responsibility to meet these.
5. A climate of opinion favorable to the development of professional programs as these might vary from traditional programs of graduate work.
6. Presumption of some ability to contribute to financial support of this additional field of graduate study.
7. Availability of services such as counseling, housing, recreation, and the like, in addition to instruction.
8. The existence of an autonomous school or college of nursing with a sound basic baccalaureate program, preferably already nationally accredited.
9. Opportunity for ample clinical experience of high quality, including public health, and opportunity to study well individuals in addition to those who are ill.

Now six universities are in the process of developing advanced programs in nursing education leading to the master's degree.

This approach has had several important beneficial results. Among them are the following:

1. It has brought together representatives of the profession of nursing education and of the universities to explore jointly the needs in this field and to identify the resources of the region to meet the needs.
2. It has been a means of securing financial support from two of the major foundations. The Kellogg Foundation has provided approximately a half million dollars for a five-year period to underwrite the cost of instruction. The Commonwealth Fund has provided scholarship and fellowship funds in the amount of twelve thousand dollars annually to each of the six schools. This fellowship program is now in its third year. The Commonwealth Fund has also provided from twenty-three to twenty-four thousand dollars for each of the first two years for the conduct of seminars. It is anticipated that this support will be continued on a reduced scale with the idea that the seminars should eventually become self-supporting.

With reference to evaluation, one further fact needs to be noted. A special evaluation committee has been appointed to make an annual appraisal of the progress of the seminars and of the education project. Also, forms have been developed for recording the progress of students, and plans are underway to follow up for a period of several years individuals who complete the programs and enter positions in the field of nursing education. Finally, a plan is being developed for a full evaluation of the project at the end of a five-year period, the term of the Kellogg grant. This is an excellent illustration of the development of a project in which evaluation is basic.

Now may I turn to a second illustration of a somewhat different nature. I refer to the Southern Regional Education Board's project in mental health. In the fall of 1953 the Southern Governors requested the Southern Regional Education Board to study the needs and resources in training and research in mental health programs of the sixteen states. This project grew out of a report on the training and research in state mental health programs prepared by the Council of State Governments. The directive to the Southern Regional Education Board from the Governors provided that the Board:

> Begin an immediate survey of facilities for the training of psychiatric personnel in the South, and that it report to this Conference those institutions best qualified to take additional students in the psychiatric disciplines from states which have no such training facilities.

That the Board also initiate a survey of institutions doing mental health research in the South, and that it recommend to this Conference those institutions capable of being enlarged to do addtional research.

That upon completion of the above surveys by the Board, but in any case not later than July 30, 1954, a Southern Regional Mental Health Conference be held to discuss the surveys and to draw up interstate contracts in mental health research and training.

That in the interim the individual states make official surveys of their training and research facilities with particular emphasis on raising the mental institutions in every state to the level of residency or affiliate accreditation, and that results of these surveys be presented to the 1954 Regional Mental Health Conference.

That the Southern Regional Education Board be requested to report the results of this study and any action taken to the 1954 Southern Governors' Conference.

A commission composed of fourteen distinguished persons representing professional or political interests in mental health was appointed to define the objectives and to determine the scope of the project. State committees appointed by the governors of the several states gathered data in their respective states on mental health needs in relation to resources and facilities available, and drafted specific recommendations. A director of the total project and his staff put together the findings of the state committees and added data gathered on a regional basis in order to get a total picture of mental health needs and the present resources to provide for these needs in the region.

Five technical panels were also set up to give attention to special problems not covered by state surveys. Each panel presented its findings to the commission. The Conference on Mental Health Training and Research, attended by some 170 persons, held in the summer of 1954 received the findings and recommendations of the state committees, the technical panels, and the staff. Finally, in September 1954 the findings and recommendations previously developed were presented to 56 legislators from the Southern Regional Compact states for their critical study. They endorsed the recommendations and commended them for consideration by the Governors of the sixteen states concerned. Now these recommendations are being put into effect.

If time permitted I could review the procedures employed in assessing needs and resources to provide for them in many other fields. The plan of organization varies from one project to another, but generally the process includes the identification of the problems, the gathering of pertinent data, the interpretation of the data, the formulation of appropriate recommendations, and the development of plans for giving effect to the recommendations. Not infrequently, as in the case of nursing education, these plans also require the identification of institutions in the region that are most competent to undertake specific research or training programs.

The second kind of evaluation with which this paper is especially concerned pertains to research that will aid in the improvement of instruction, organization, and administration in higher education. In 1951 the Board invited a group of educators from the Southern region and several consultants from outside the region to consider the problems of institutional self-evaluation. The net outcome of this and subsequent conferences was the publication of a manual entitled, ''Improving Graduate Education: A Guide to Institutional Self-Evaluation.''

The impact of this approach was relatively limited. Now the Board is moving in the direction of more directly promoting research and experimentation in higher education. This is another approach to protecting and improving the quality of higher education. The Board foresees that the quality of higher education will be threatened by the impending tidal wave of students and believes that now is the time to examine the validity of the premises on which the colleges and universities operate. This approach points in three directions: intra-institutional, statewide inter-institutional, and regional studies.

The nature of needed studies and experiments is suggested by the major problems confronting higher education in the United States, namely,

1. Impending critical shortages of both the supply and quality of college teachers. The need therefore, to hold and to make more effective use of superior teachers.
2. The growing complexity of the curriculum reflected in increased offerings of fractionated and specialized courses.
3. The employment of conventionalized methods of instruction.
4. The cumbersome plans of institutional organization and management.
5. The unnecessary and uneconomical duplication of programs among state colleges and universities and the duplication of expensive and highly specialized programs within the region.

The delineation of these problems implies the

the urgent need for carefully controlled research and experimentation in such areas as the selection, preparation, and use of faculty members, curriculum re-organization, methods of teaching, intra-institutional and inter-institutional organization and administration. To be more specific, what combination of qualities and procedures makes for superior teaching, what is a reasonable faculty load, what use can be made of television and other audio-visual aids to extend the compass of superior teachers, how can economies and improved effectiveness be achieved through curriculum organization, what plans of space construction and utilization contribute to increased effectiveness and economy, what are the merits and limitations of various plans of institutional and statewide administration? With reference to the research and experimentation of the kind just outlined, the Board will serve as the facilitating agent. It will stimulate institutions and state boards to explore these problems and, on request, will aid in designing appropriate research and experimentation. The Board itself will undertake projects in this area only as it is requested to do so by the institutions and the states of the regions. The Board is still in the early stage of promoting the much needed evaluative research and experimentation in this area.

The anticipated outcome of the Board's projects is generally a program of action. But in order that the Board may achieve its predetermined purposes it must identify and gear itself into the dynamic and powerful elements of the culture of the states and of the region. Especially important is the appraisal of these cultural components as to their relative importance, their favorable or antagonistic attitudes, and the possibility of melding them into a régional pattern.

Two of the most important of these elements in the Southern region are state governments and the colleges and universities. How to appraise the relative dynamics of these two elements and how to fuse or coordinate them in the interest of the social and economic development of the South is one of the Board's major problems. Two main methods of assessing these cultural forces have been employed: viz., the judgment of educational and political leaders and practical experience; neither of them scientific in the strict sense of the term. To gear into the dynamics of administrative personnel in state governments has presented no great problem because the Southern Regional Education Board is the agent of the states and was created by the governors of the compact states. But the election of a new governor, one who had nothing to do with the creation of the Board and who knows little about its functions, may greatly impair the dynamic power of the group. Such a situation immediately requires the identification of avenues through which the governor may be given an understanding of the Board's program and may become favorably disposed to it.

To tie the Board's program into the legislative bodies of the states is another matter. One approach that the Board has developed is the Legislative Work Conference—a conference in which about 60 legislators from the compact states take part each year. Consideration of regional problems in higher education by this group has been an important means of creating common interests and understandings that transcend state boundaries.

An appraisal of these procedures has indicated, however, that the higher institutions through their presidents and boards of control must somehow be brought into the picture more directly. This is especially difficult because of the variety of plans for the administration of state-supported higher institutions. One idea that is being tried experimentally is to request the presidents of universities who are members of the Southern Regional Education Board to invite members of their boards of control to attend the Legislative Work Conferences. Whether this will be a means of fusing another element into the regional pattern the Board is attempting to develop remains to be seen. If the plan appears to be sufficiently promising it may be extended to include members of boards of control of other universities in the region.

It seems quite clear that as time goes on it will be necessary to identify other dynamic elements in the culture pattern of the South and to attempt to marshal their forces in support of the Board's program. There may be better methods for doing this than those which the Board has relied upon but so far those better methods have not come to light.

It is important to note that the Board's continuing assessment of higher educational needs in the South and the development of programs to provide for identified needs currently involves about twenty committees and commissions with a total membership of nearly 300 professional and lay persons. This number does not include participants in the Legislative Work Conferences that are held from time to time. Significant also are some of the outcomes of these exploratory and evaluative activities; for example, contract programs among selected universities of the region in the fields of dentistry, medicine, social work, and veterinary medicine now enroll about 1000 students to whom the advantages of education in these professional fields were not available in their respective states; and memoranda of agreement among universities offering specialized programs in city planning, forestry, marine sciences, nursing, petroleum sciences, and pulp and paper have already been adopted. Exploratory studies still underway will in all probability lead to additional memoranda. From these activities have also come publications and consultations designed to aid institutions and professional groups in identifying problems and in finding solutions to them.

I have sketched very briefly three types of evaluation that are involved in the development of regional programs of higher education—assessing regional needs and resources to meet those needs; stimulating and aiding institutions to assess by research and experimentation the basic premises underlying instruction, curriculum organization, and administration; and assessing the dynamic elements of a regional culture pattern that may be drawn upon to put programs into effect. I have also indicated in a few words some of the achievements in a regional approach to the improvement of higher education. These illustrations will suffice to support the statement I made at the beginning, namely, evaluation plays an important role in the development of regional programs of higher education.

THE ROLE OF SIXTEEN MILLIMETER MOTION PICTURES AND PROJECTED STILL PICTURES IN SCIENCE UNIT VOCABULARY LEARNING AT GRADES FIVE, SIX, AND SEVEN

LOUIS ROMANO
University of Wisconsin

Definition of the Problem

THIS STUDY is an attempt to identify the role of sixteen millimeter motion pictures and projected still pictures in science unit vocabulary learnings in grades five, six, and seven. This study specifically attempts to find possible answers to these questions: (1) Do children learn more of the vocabulary of science units when motion picture films and projected still pictures are added to the use of other audio-visual materials? And (2) do children retain more of the vocabulary of a science unit when motion picture films and projected still pictures are added to the use of other audio-visual materials?

The subjects in this study were boys and girls from two fifth, two sixth, and two seventh grade groups in the public schools of Shorewood and Whitefish Bay, Wisconsin. These groups were rotated so that each served in turn as an experimental and a control group. Blackboards, bulletin boards, charts, models, flat pictures, and field trips were used both in the control and experimental situations, but motion pictures and projected still pictures (filmstrips 2 × 2 and 3-1/4 × 4 slides, and pictures used with the opaque projector) were used only in the experimental situations and served as the experimental factors. Vocabulary test results, supplemented by the reactions of teachers and students, provided the means of judging the contributions of the experimental factors.

Significance of the Problem

It should be pointed out that this study does not attempt to determine the effectiveness of a single audio-visual material over other visual materials such as slides, demonstrations, maps, and the like, or the more traditional classroom procedures using verbal instruction by means of textbooks or supplementary reading. Instead, it attempts to determine the effectiveness of an audio-visual utilization in which several audio-visual materials are employed over an audio-visual utilization in which a limited number of audio-visual materials are used. Dale[1]* states that experimentors often neglect to establish normal schoolroom procedures in their investigation which may result in lessening the value of their data. The procedures in this study seem to follow a realistic classroom situation. Many schools use some audio-visual materials, but according to various surveys many schools do not have an audio-visual program which includes the use of several audio-visual materials in the instructional program. If the findings of this study should point out that better learning is achieved when certain audio-visual materials are employed throughout a unit of study, it might suggest that these materials be included as an integral part of the instructional program.

In previous experiments, emphasis was placed on the values of films for imparting factual information.[2] This study may tend to give evidence that the use of audio-visual materials in the instructional program contributes to better learning experiences. Educators cannot refrain from producing research that points out the effectiveness of an instructional tool. This is especially true because many magazine articles and newspaper stories point out the public's concern for the teaching of the fundamental skills. The findings of this study may be used by educators in informing the public that the use of audio-visual materials helps children in gaining the vocabulary comprehension which is vital to any basic skills program.

Classroom teachers are concerned about the ability of boys and girls to acquire and retain the vocabulary taught during an instructional unit of study. The findings of this study may point out that the vocabulary of a unit of study may be learned when an aggregate of audio-vis-

*All footnotes will be found at the end of this article.

ual materials as defined for the experimental group is used in the teaching-learning situation. Learning the vocabulary of the unit of study for the purpose of knowing words *per se* is not as important as knowing the words which will be used to grasp the important concepts, understandings, or generalizations of the unit of study. Acquiring the vocabulary in the unit of study might help children grasp the meanings of the generalizations more readily.

Limitations of the Study

In spite of careful planning by the participating teachers and the investigator, this study is not without limitations.

Probably the most obvious limitation might seem to be the small number of students from which the data was obtained. This study was limited to the collection of data in the present situations, and is best described as a descriptive study.[3]

Another limitation grows out of the fact that this study is limited to boys and girls in the middle to upper class cultural backgrounds. Both communities, Shorewood and Whitefish Bay, enjoy many cultural and social advantages which one might not find in a large city with its heterogeneous population.

A further limitation is found in the vocabulary tests. They were primarily developed as recognition tests and they do not in any way tell how children should use these words.

The vocabulary words in the objective tests were obtained from the text materials included in the science units as taught in the fifth, sixth, and seventh grades. A more comprehensive picture would have included determining whether or not children acquire and retain the vocabulary in other subject areas such as social studies and the language arts. To a certain extent vocabulary used in the science unit of study is also included in other areas of study, but this is limited.

A limited number of appropriate films and filmstrips for various units of study in elementary science made it difficult for the participating teachers to obtain the best audio-visual material for the instructional program.

Procedures Employed in this Study

Rotation Technique. The rotation-group technique of experimentation was used in order to minimize the effect of several uncontrollable factors such as initiative, industry, or study habits. The rotation method ". . . . involves an exchange for the groups at intervals, in terms of the procedures followed."[4] This technique was used with both the classes and the teachers. During each unit of study for each grade level, one class was designated as the "control" group while the other class was designated as the "experimental" group. The "control" group was the group which used only the following audio-visual materials: bulletin boards, blackboards, charts, models, flat pictures, and field trips. The "experimental" group not only used the materials listed for the control group, but also used sixteen millimeter motion picture films and projected still pictures.

Selection of Groups Participating. Equal numbers of children in the fifth and sixth grades from the Lake Bluff School of Shorewood, Wisconsin, and from the seventh grade from Cumberland School of Whitefish Bay, Wisconsin, were included. These groups were equated within feasible limits on the basis of I. Q. scores and standardized reading vocabulary test scores.

The Selection of the Vocabulary Words and the Construction of Tests. Prior to the experiment, the investigator reviewed all of the science textbooks from the third grade level through the ninth grade level to assemble the vocabulary essential to the understanding of the core ideas to be developed within a unit of study. In addition to looking through the available textbooks, the investigator reviewed each article in the *World Book Encyclopedia*[5] that pertained to a particular unit of study. A review by the investigator of all the available textbooks in each of the participating schools from the third grade level through the ninth grade level, was designed to provide a range of science words from the third through the ninth grade level.

These lists of words were then carefully reviewed by a committee made up of teachers participating in this study. Also, teachers who were especially interested in the teaching of science and had had several years of experience in this field, were asked to check this list of words carefully. After the vocabulary lists were approved by this committee, test items were developed which included all of the words in the list. In each unit approximately 250 words were recorded. This made possible the construction of a 50-item vocabulary test. Wherever possible a word meaning nearly the same as the key word was obtained from the list of 250 words. For example, part of the vocabulary from the sixth grade unit of study, *Sound*, includes the following:

reception
range
transmit
amplitude
supersonic

loudness
spread
mute
intensity
audio

decrease
velocity
speed
amplify

From the above list, the committee of teachers selected and designated the word "velocity" as a key word. The test item then included five different words, one being the correct answer. Whenever possible a word which had the same, or nearly the same meaning was obtained from the list of 250 words. This word was considered the correct response.

The tests used as instruments of measure in this study were of the multiple-choice objective type.

Validity and Reliability of the Test Items. A high degree of curricular validity could be assumed since the vocabulary was closely correlated with the curricular content of each unit of study.

Reliability was estimated by administering the vocabulary tests to groups not participating in this study. The split-half method was used. This necessitated obtaining separate scores on odd and even-numbered items. A coefficient of correlation was obtained between these two scores by the product-moment method. By means of the Spearman-Brown formula, an estimate of the reliability of a test as long as the two halves combined was obtained.

Administration of the Vocabulary Tests. Prior to teaching the science units, tests for each grade level were given to the control and experimental groups. These tests were administered two hours apart so as to reduce practice effect. The scores are designated as the "pre-test scores". At the termination of each unit of study which took five weeks, "final tests" were given. Six months after the "final tests", the participants were given the same test. These scores are designated as "re-test scores".

To insure objectivity, teachers involved in this study did not check these papers. Objective answer sheets were worked out, and all tests were scored on the basis of these answer sheets.

The Selection of the Science Units of Study. A committee of teachers selected two units for each grade level based on the following criteria: (1) availability of audio-visual materials related to the various units of study, and (2) extent of previous experiences in the particular unit of study. Specific problems to be covered in each unit of study were defined by the teachers so as to insure some degree of uniformity in both experimental and control groups.

The Selection of Audio-Visual Materials for the Experimental Group. The criteria determined by the participating teachers included: (1) Does the audio-visual material contribute to the objectives of the unit of study? (2) Does the audio-visual material suit the experiences, intellectual maturity, and grade level of the students? (3) Is the audio-visual material accurate and authentic? (4) Is the audio-visual material significant? Any audio-visual materials previewed that did not meet the above criteria were not used in this study.

The procedure for showing audio-visual materials was carefully planned by the group. Such procedures were demonstrated in an in-service meeting.

Procedure in the Selection of Pupil and Teacher Responses. Subjective evidence was gathered on the reactions of pupils and teachers to the utilization of audio-visual materials in the classroom. Specifically, teachers were asked to respond in writing to the following question: "What reactions do you have concerning teaching a unit of study in which a limited number of audio-visual materials are used as contrasted with a situation in which several audio-visual materials are used." The reports were then submitted at the termination of the ten-week experimental period.

The boys and girls were asked to react to the following question: "What reactions do you have concerning the first unit of study in science and the second unit of study?" Responses to this question were recorded immediately by the participating teachers, or, when possible, a stenographer was employed.

Presentation of Findings and Interpretation: Test Results

The findings for the fifth, sixth, and seventh grade groups using the rotation-group technique are summarized in Table I.

Table I points out that the fifth grade classes in the unit on Electricity had a mean score of 7.24 words on the pre-test for the experimental group, and a mean score of 11.45 for the control group. In the final test scores given five weeks after the pre-test, the experimental group scored a mean of 23.62 words while the control group scored a mean of 21.93 words. In this particular unit the experimental group had a mean gain of 16.38 while the control group had a mean gain of 10.48. The experimental group in the unit on Electricity had a mean gain of 5.90 words over the control group.

In the unit of study, Rocks, the experimental group had a mean pre-test score of 15.06 while the control group had a mean pre-test score of 17.40. In the final tests the mean scores were 29.38 for the experimental group, and 21.10 for the control group. This represents a mean gain of 14.31 for the experimental group, and a mean gain of 3.76 for the control group. The experimental group then had a mean gain of 10.55 words over the control group.

The same table summarizes the findings for the sixth grade unit on Astronomy. In this unit of study the experimental group had a mean pre-test score of 15.68 while the control group had a mean pre-test score of 14.00. The final test for the unit on Astronomy produced the following test scores: experimental group a mean score of 34.84 and the control group a mean score of 21.72. The experimental group had a mean gain of 19.16 while the control group had a mean gain of 7.72. The experimental

TABLE I

SUMMARY OF MEAN SCORES ACHIEVED IN PRE-TESTS, FINAL TESTS AND GAINS FOR BOTH EXPERIMENTAL AND CONTROL GROUPS IN ALL UNITS OF STUDY

Grade	Unit of Study	Pre-Test Mean Scores		Final Test Mean Scores		Mean Gains*	
		Control Group	Experimental Group	Control Group	Experimental Group	Control Group	Experimental Group
5	Electricity	11.45	7.24	21.93	23.62	10.48	16.38
5	Rocks	17.40	15.06	21.10	29.38	3.76	14.31
6	Astronomy	14.00	15.68	21.72	34.84	7.72	19.16
6	Sound	15.00	13.88	21.52	30.72	6.45	16.84
7	Air	19.80	22.68	24.16	35.68	4.36	13.00
7	Soil	21.68	22.64	25.48	34.20	3.84	11.56

*Final Test Score Minus Pre-Test Score

group then had a mean gain of 11.44 words over the control group.

In the unit of study, *Sound*, the experimental group had a mean pre-test score of 13.88 while the control group had a mean pre-test score of 15.08. In the final tests the mean scores were 30.72 for the experimental group and 21.52 for the control group. This represents a mean gain of 16.84 for the experimental group, and a mean gain of 6.45 for the control group. The experimental group scored 10.39 more words per child over the control group.

In the seventh grade on *Air* the experimental group had a mean pre-test score of 22.68 and a mean final test score of 35.68. This is a mean gain of 13.00. The control group had a mean pre-test score of 19.80, and a mean final test score of 24.16. The control group showed a mean gain of 4.36. This shows that the experimental group had a gain of 8.64 over the control group.

In the seventh grade unit of study on *Soil* the experimental group had a mean pre-test score of 22.68, and a mean final test score of 34.20. This is a mean gain of 11.56. The control group instead had a mean pre-test score of 21.68, and a mean final test score of 25.48. This represents a gain of 3.84 for the control group. Comparing the mean gains, the experimental group increased 7.72 over the control group.

Table II summarizes the percentage of gain achieved by the experimental group over the control group in the amount of vocabulary acquired by the boys and girls in these groups.

What may be interpreted from the data presented in Table I? All experimental groups using the motion picture films and projected still pictures had a range of mean scores from 6.90 to 11.44 words gained over the control groups. This indicated a large increase in words gained over the control groups. Thus it may be concluded that the children of this particular population acquired to a greater extent the vocabulary in a unit of study when motion picture films and projected still pictures were introduced.

Six months following the completion of the unit of study the boys and girls were given vocabulary tests on the particular unit so as to determine the amount of vocabulary retained over this period of time. Table III summarizes the results achieved in the re-test vocabulary tests.

A close examination of the re-test and final test scores shows that re-test scores of the experimental group in the *Rock* and *Astronomy* units of study for the experimental group and for all the units of study of the control groups, the re-test scores were higher than the final test scores. There was a loss in retention for the experimental group in only two of the units of study. The following may indicate reasons for this increase in scores: (1) the science program did not terminate following the experimental period. All classes continued with the science instructional program, and throughout the instructional program classes utilized many audio-visual materials including films and projected still pictures. An examination of the units of study and audio-visual materials used following the experimental period show that some of the vocabulary words included in the vocabulary tests were used. This may account for the gains in the re-test vocabulary scores for both the control and experimental groups. (2) During the period between the pre-test and the final test both science and social studies were taught as part of the daily instructional program. An attempt was made so as not to include social studies units which were closely allied or related to the science units of study. This procedure may eliminate to a certain extent the duplication of terms or vocabulary which may be included in the vocabulary tests. Following the final tests the social studies and science programs may be correlated, but in any case both classes received the same or similar activities.

An analysis was done so as to check on the possibilities that the experimental group gains could be accounted for by the presence of key words in the media to which the control group may not have had access.

The sixth grade unit of study on *Astronomy* was chosen to study carefully the vocabulary included in the audio-visual materials which was identical to the vocabulary in the test. Narrations of films and filmstrips were obtained. Many duplications of words were found, but only 138 different words were included in audio-visual materials as compared to 250 words in the vocabulary tests. This is 55% of the total number of words included in the *Astronomy* vocabulary tests. Of the 138 words only 14 were key words or only 28% of the 50 key words included in the test.

An analysis of the data show that the experimental group and the control group had approximately the same means in the pre-test, 4.84 for the experimental group, and 4.08 for the control group. In the final test, the experimental group scored a mean of 11.50 while the control group scored a mean of 6.84. This is a mean gain of 4.66 words over the control group. The total gains for the experimental group in this unit of study was 19.16 on the entire test, whereas the gains on the key words was only 6.66.

If the gains from both the experimental and control group in the key words were subtracted from the mean gains achieved in the whole test, then the control group would have a mean gain of 4.96 while the experimental group would have a mean gain of 12.50. These data show that the gains achieved by the experimental groups were not gained primarily from the showing of films and projected still pictures. Stated in another way, the gains achieved by the experimental

TABLE II

PERCENTAGE OF GAIN ACHIEVED BY CONTROL GROUP COMPARED TO THE EXPERIMENTAL GROUP IN ALL UNITS OF STUDY IN GRADES FIVE, SIX, AND SEVEN

Grade Level	Unit of Study	Actual Mean Gain Control I	 Experimental II	Percentage of Gain Achieved by Control Group Compared to Experimental Group
5	Electricity	10.48	16.38	63.9%
5	Rocks	3.76	14.31	26.2
6	Astronomy	7.72	19.16	40.3
6	Sound	6.45	16.84	38.3
7	Air	4.36	13.00	33.5
7	Soil	3.84	11.56	33.0

TABLE III

SUMMARY OF MEAN RE-TEST SCORES AND MEAN GAINS FOR BOTH EXPERIMENTAL AND CONTROL GROUPS IN ALL UNITS OF STUDY

Grade	Unit of Study	Re-Test Scores-Means		Mean Gains Re-Test Over Final Test	
		Control Group	Experimental Group	Control Group	Experimental Group
5	Electricity	25.41	27.82	3.48	4.20
5	Rocks	22.62	27.97	1.52	-1.41
6	Astronomy	26.52	33.08	4.80	-1.76
6	Sound	25.08	31.68	3.56	0.96
7	Air	31.64	38.84	7.48	3.16
7	Soil	31.80	39.16	6.32	4.96

group were acquired from the total instructional program rather than from the use of films and projected still pictures. Therefore, the gains achieved by the experimental group cannot be accounted for by the vocabulary included in the films and projected still pictures.

Is There a Relationship Between the I.Q. and the Child's Gains in the Vocabulary Tests in the Experimental Group?

The Pearson-Product-Moment Test for correlation was found between the child's I.Q. score and the gains he achieved in the vocabulary tests. The vocabulary test gains were derived by subtracting the pre-test score from the final test score. In each of the units of study, the correlations obtained are shown in Table IV.

In the units of study on Astronomy, Rocks, and Air, there is no relationship while in the remaining units of study, Electricity, Sound, and Soil, there is a slight relationship between I.Q. scores and gains achieved in the vocabulary tests. In conclusion, no definite relationship exists as to I.Q. scores and gains achieved in the vocabulary tests in any of the units of study on the fifth, sixth, and seventh grades. The foregoing conclusion also means that the bright, average and slow child profited somewhat the same in this study, or stated in another way, the gains achieved in the vocabulary tests do not necessarily depend upon the intellectual level of the individual.

Is There a Relationship Between the Standardized Reading Vocabulary Scores and the Gains Achieved in the Vocabulary Tests by the Experimental Groups?

To answer this question, the Pearson-Product-Moment Test of Correlation was obtained between the standardized reading scores and the gains achieved in the vocabulary tests in this study.

The data shows that in only two units of study (Rocks and Air) there is a definite correlation which exists between standardized reading vocabulary scores and the vocabulary test scores in this study. In the remaining four units of study there is only a slight relationship.

Pupil and Teacher Reactions to Audio-Visual Materials. The reactions of pupils and teachers represent the subjective evaluation of the use of various audio-visual materials in the classroom. All of the teachers in the study point out the intrinsic value of the use of audio-visual materials in making for a more effective teaching-learning situation. The following excerpts from the teacher's written reactions emphasizes this point:

"Audio-Visual devices are necessary experiences for the children in order to gain the fullest appreciation of the concepts presented in any given unit of study."

"....greater enthusiasm for extensive and intensive study by the pupils."

"....projective aids developed a greater interest and better understanding."

"....more voluntary suggestions for additional activities were forthcoming from students in the experimental group."

Some of the students' reactions are as follows:

"Films are better because things are illustrated."

"We learned more because we could see what was meant instead of reading about it or hearing it alone."

"Astronomy filmstrips are more interesting. They help me to get a better idea of what the book said."

"I'll never forget how the layers of earth folded over, I just couldn't understand that when we talked about it in class, but I get it right away from the film."

The reports from both teachers and pupils in a situation in which a limited number of audio-visual materials were employed is a negative one. An enthusiastic learning situation seemed to develop in the classroom situation in which several audio-visual methods were used.

Conclusions

1. All experimental groups using the motion picture films and projected still pictures had a large increase in vocabulary gained over the control groups in all of the units of study.
2. Re-test scores of the experimental group in the Rock and Astronomy units of study were lower than the final test scores while in all the other units of study for the experimental group and for all the units of the control groups, the re-test scores were higher than the final test scores. There was a loss in retention for the experimental group in only two of the units of study.
3. No definite relationship seems to exist as to I.Q. scores and gains achieved in the vocabulary tests in any of the units of study on the fifth, sixth, and seventh grades.
4. In only two units of study (Rocks and Air) there is a definite correlation which exists

TABLE IV

CORRELATIONS BETWEEN I. Q. SCORES AND GAINS ACHIEVED IN THE VOCABULARY TESTS BY THE EXPERIMENTAL GROUPS IN THE SCIENCE UNITS, GRADES FIVE, SIX, AND SEVEN

Grade	Unit of Study	Correlation
5	Electricity	+.38
5	Rocks	-.01
6	Astronomy	-.15
6	Sound	+.21
7	Soil	+.33
7	Air	+.06

TABLE V

PEARSON PRODUCT-MOMENT TEST FOR CORRELATION BETWEEN READING VOCABULARY SCORES ON STANDARDIZED TESTS, AND GAINS ACHIEVED IN THE VOCABULARY TESTS FOR ALL UNITS OF STUDY

Grade	Unit of Study	Correlation
5	Electricity	+.18
5	Rocks	+.54
6	Astronomy	-.21
6	Sound	+.09
7	Soil	+.18
7	Air	+.69

between standardized reading vocabulary scores and the vocabulary test scores in this study. In the remaining four units there is only a slight relationship.

5. All of the teachers in this study point out the intrinsic value of the use of audio-visual materials in making for a more effective teaching-learning situation.
6. The boys and girls participating in this study point out that an instructional program which uses many audio-visual materials enhances the learning situation.

Implications

The implications of the findings of this study are presented in relation to the role of the sixteen millimeter film and projected pictures in science unit vocabulary learnings in grades five, six, and seven.

The fact that all of the experimental groups in grades five, six, and seven using sixteen millimeter films and projected still pictures had a large increase in vocabulary learnings as compared to the control groups in all of the units of study indicates that these materials should be given serious consideration in a science program aimed at vocabulary growth. It must be repeated that the experimental groups were compared not with a situation in which meager number of audio-visual materials were used, but rather with one in which several audio-visual materials were used except for the sixteen millimeter film and projected still pictures. This suggests that the instructional program should not be limited to the sixteen millimeter film and projected still pictures, but should include a multiple number of audio-visual materials, i.e., blackboards, bulletin boards, charts, models, flat pictures, and field trips. All of these materials carefully chosen and properly utilized should enhance the teaching-learning situation.

One of the major objectives in the teaching of science is the development of important understandings or concepts. These concepts are expressed through word symbols, that is, vocabulary. The ability to interpret word symbols may play an essential part in aiding comprehension and understanding of these concepts in science. This study may have important implications for the enrichment of the child's basic science concepts.

Teacher-training institutions have an obligation to train prospective teachers in the use of a number of audio-visual materials in the instructional program. Too often, beginning teachers limit their instructional program to the use of one or two audio-visual materials rather than a number of them. If teacher-training institutions fail to help prospective teachers, then it is the responsibility of the school administrator or a designated person, such as the audio-visual chairman, to assist teachers in learning how to use the various audio-visual materials. Learning how to operate the various projectors is important, but even more important is learning how to use these materials in the instructional program.

A program of instruction which includes a number of audio-visual materials demands budgetary provisions for their purchase. School board members must be enlightened by their superintendent as to the intrinsic value of using audio-visual materials in the instructional program. All staff members should be encouraged through salary provisions to attend summer school and enroll in audio-visual courses. Arrangements should be made for teachers to attend audio-visual institutes or workshops in which a broad approach to instructional materials is used.

The acquisition of vocabulary, one of the important skills in reading, is a prerequisite to adequate comprehension and interpretation in reading. The utilization of audio-visual materials enhances the acquisition and retention of vocabulary, and they should be included in the instructional program.

FOOTNOTES

1. Edgar Dale, Fannie W. Dunn, Charles F. Hoban Jr., and Etta Schneider, Motion Pictures in Education, The H. W. Wilson Company, New York, New York, 1938, p. 319.
2. Ibid., p. 325.
3. Arvil S. Barr, Robert A. Davis, and Palmer O. Johnson, Educational Research and Appraisal, J. B. Lippincott and Company, New York, New York, 1953, p. 310.
4. Carter V. Good and Douglas E. Scates, Methods of Research, Appleton-Century-Crofts, Inc., New York, New York, 1954, pp. 706-707.
5. The World Book Encyclopedia, Field Enterprises, Inc., 1952.

RELATION BETWEEN ENROLLMENT OF HIGH SCHOOLS FROM WHICH STUDENTS GRADUATED AND ACADEMIC ACHIEVEMENT OF AGRICULTURAL STUDENTS, A. AND M. COLLEGE OF TEXAS*

JOHN R. BERTRAND**
A. and M. College of Texas

THE AGRICULTURAL and Mechanical College of Texas, as a state supported educational institution, admits all white male students graduating with, at least, fifteen acceptable units from an accredited secondary school. One result of this unselective admission is the enrollment of a number of students unprepared to do college level work.[1] A higher proportion of these unprepared students enroll in agricultural curricula than in arts and sciences or in engineering curricula. Student loss is heavy among these unprepared students. This article reports the results of a study of the relation between enrollment of high schools from which students graduated, scores on aptitudes and achievement tests, and academic achievement of agricultural students.

Previous Research

During the past thirty years, many studies have been made on forecasting college achievement. Some of these, with findings much the same as for other studies, have been on students of agriculture. High school grades and scores on various psychological tests have been the predictive measures used most often and grades given by the college teachers nearly always have been the criterion of achievement. Scores on achievement tests administered at the time of college entry have been used in only a few cases as predictive measures of scholastic achievement of agricultural students. Previously, little attention has been given to investigating possible relations between enrollment of high schools from which students graduated, scores on aptitude and achievement tests and academic achievement.

Sources of Information

The over-all study of which this is a part was concerned with 637 non-transfer students who first enrolled as freshmen in agricultural curricula at the Agricultural and Mechanical College of Texas in 1946, 1947 and 1948, when because of overcrowded conditions on the main campus an annex was established to care for students during their freshmen year. An additional 195 students, enrolling at the same time, were not studied either because of non-comparability of course work or because of lack of essential information. Enrollments of high schools from which students graduated were available for 621 of the 637 students with which most of the over-all study was concerned. For purposes of analyses, five enrollment categories were set up as follows: (1) under 150; (2) 150-249; (3) 250-499; (4) 500-999; and (5) 1000 and over.

The measures used were: high school average grade, the three scores (quantitative, linguistic and total) from the A.C.E. Psychological Examination (1946 edition), and scores on the Co-operative Achievement tests in mechanics of expression, reading, mathematics and science. The criteria of achievement which were used were: grade point ratio[2] for all courses taken in English, in natural sciences, and in agriculture by the end of the second year, and the grade point ratio for the total of all courses taken by the end of the second year.

In order to determine the relationship of en-

---

* Adapted from writer's unpublished Ph.D. thesis, "Relation of Selected Social and Psychological Factors to Academic Achievement of Students in the School of Agriculture, Agricultural and Mechanical College of Texas," Cornell University, 1950.

** Present address: College of Agriculture, University of Nevada, Reno, Nevada.

[1] In 1950, the A. and M. College of Texas established its Basic Division for the purpose of providing close supervision and assistance to all students during their period of transition from high school to college.

[2] Grade point ratio is determined at the A. and M. College of Texas by dividing the total hours for which a student has been registered, whether passed or failed, into the total grade points earned. For each semester hour of credit which a course carries, the following grade point scale is used: A = 3; B = 2; C = 1; D = 0; and F = 0. However, for the purpose of this study, a negative 1 was assigned for grades of F.

TABLE I

RANGE OF GROSS SCORES, A.C.E. PSYCHOLOGICAL EXAMINATION, LOCAL NORMS* QUARTILE CATEGORIES

Quartiles	Gross Scores
Lowest Quartile	Under 86
Second Quartile	86 - 100
Third Quartile	101 - 114
Highest Quartile	115 and over

*Unpublished data from the Office of the Registrar. N equals 1126 entering, non-veteran students from the 1946-47 freshman class at the Agricultural and Mechanical College of Texas.

TABLE II

DISTRIBUTION OF 621 STUDENTS* BY ENROLLMENT OF HIGH SCHOOLS FROM WHICH STUDENTS GRADUATED AND A.C.E. PSYCHOLOGICAL EXAMINATION QUARTILE CATEGORIES**

Enrollment of high schools from which students graduated	Total no. of Students	In ACE lowest quartile	In ACE second quartile	In ACE third quartile	In ACE highest quartile
		Number of Students			
Under 150	154	87	43	17	7
150 - 249	133	64	35	17	17
250 - 499	159	74	30	35	20
500 - 999	60	22	17	16	5
1000 and over	115	28	41	32	14
Total	621	275	166	117	63
	% of total	Percent of total in each enrollment group			
Under 150	24.80	56.49	27.92	11.04	4.55
150 - 249	21.42	48.12	26.32	12.78	12.78
250 - 499	25.60	46.54	18.87	22.01	12.58
500 - 999	9.66	36.67	28.33	26.67	8.33
1000 and over	18.52	24.35	35.65	27.83	12.17
Total	100.00	44.28	26.73	18.84	10.15

---

* Enrollment of high schools was unknown for 16 of the 637 students with which most of this study is concerned.

**Agricultural Curricula, A. and M. College of Texas, 1946-48 entering classes at Annex.

rollment of high schools from which students graduated to achievement, when the effect of aptitude is controlled, much of the analyses have been accomplished with tables which controlled the effects of aptitude as measured by gross scores on the A. C. E. Psychological Examination. Quartile categories based upon local norms (Table I) have been the means for controlling this score for purposes of comparisons.

Findings

Students graduating from high schools with enrollments of under 150 constituted 24.80 percent of the total (Table II). Over one-half, 56.49 percent, of the students from this category ranked in the lowest quartile on the A. C. E. Psychological Examination while a total of 84.41 percent ranked in the lowest and in the second quartiles combined. Only 4.55 percent of the students ranked in the upper quartile. Students graduating from the large high schools with enrollments of 1000 students or over were 18.52 percent of the total. About one-fourth, 24.35 percent, of students in this category ranked in the lowest quartile. In the highest quartile were 12.17 percent of the students in this large school category. Forty percent of the students from this large school category ranked in the combined third and highest aptitude quartiles. It may thus be seen that for the students in this study, enrollment of high school was directly related to aptitude scores as measured by the A. C. E. Psychological Examination.

The above generalization was further verified by the mean scores on the A. C. E. Psychological Examination (Table III). Students from the small high schools had a mean gross score of 82.49. There was an increase in the aptitude gross score with each increase in size of high school. The highest average gross score for the group was 95.21. These students had graduated from the large high schools with enrollments of 1000 and over. A study of the high school average grade presented in the same table reveals that the highest average grades in high school were earned by students graduating from small high schools. The average grade for students who graduated from high schools with enrollments of under 150 was 83.81. Students from schools with enrollments of 150 to 249, inclusively, had a high school average grade of 82.44. These averages should be compared with the average of 81.50 for students from the large high schools having 1000 or more students enrolled. It appears that the low gross score for students from the small high schools was most greatly influenced by their poor showing on the linguistic portion of the A. C. E. Psychological Examination. The mean L-score for students graduating from schools with enrollments of under 150 was 46.92 as compared with a mean of 52.65 for students from the next category and of 57.40 for students from the large high schools. The range in scores was not nearly so great on the Q-score portion of this examination as for the mean L-score.

A comparison of mean scores on achievement tests, taken at the time of students' entry into college, with enrollments of the high schools from which they graduated is significant in this study. Achievement scores for mechanics of expression (Table IV) had little relationship to enrollment of high school when the effect of the A. C. E. Psychological Examination gross score was controlled by aptitude categories. It can be clearly seen that students ranking in the third and highest quartiles and who were from the small high schools had a larger mean achievement score in mechanics of expression. At the same time it is to be remembered that a much smaller percentage of students from the small schools were in these two higher quartile categories.

In the case of reading achievement (Table IV), again there seemed to be little relationship between mean scores and enrollment of high schools when aptitude was controlled. Students from the two smaller school groupings, who also ranked in the highest apptitude quartile, had a higher mean achievement score in reading than students who graduated from the large schools. It should be pointed out again that a much smaller percent of students from these small schools ranked in the highest aptitude quartile.

Achievement scores in mathematics and science at the time of students' entry into college also had little relationship, when aptitude was controlled, to the enrollment of high schools from which students graduated (Table IV).

Mean grade point ratios for all courses taken by the end of the second year appeared to be more directly related to aptitude than to enrollments of high schools from which students graduated (Table V). Students in the lowest aptitude quartile and from the small high schools, however, had a higher mean grade point ratio in all courses taken than did students from the large high schools who were in the lowest aptitude quartile. With fewer students in the highest quartile category, the significance of their mean grade point ratios was less than that of the grade point ratios of those in the lowest quartile. However, the enrollment of the high schools had little apparent uniform relationship to mean grade point ratios.

Mean grade point ratios for all courses in English taken by the end of the students' second year in college (Table V) were also more closely related to aptitude than to size of high schools as measured by enrollment. In both the lowest and the second quartiles on aptitude, students from small high schools had slightly higher ratios than

TABLE III

HIGH SCHOOL AVERAGE GRADE AND MEAN SCORES ON A. C. E. PSYCHOLOGICAL EXAMINATION FOR 621 STUDENTS BY ENROLLMENT OF HIGH SCHOOLS FROM WHICH STUDENTS GRADUATED*

Enrollment of high schools from which students graduated	High school av. grade	Mean scores on A. C. E. Psychological Examination		
		Gross score	L-score	Q-score
Under 150	83.81	82.49	48.92	33.52
150 - 249	82.44	87.50	52.85	34.58
250 - 499	81.43	89.53	52.54	37.13
500 - 999	81.23	90.48	54.45	36.08
1000 and over	81.50	95.21	57.40	37.72
Total	82.23	88.49	52.79	35.70

*Agricultural Curricula, A. and M. College of Texas, 1946-48 entering classes at Annex.

TABLE IV

MEAN SCORES ON COOPERATIVE ACHIEVEMENT TESTS FOR 621 STUDENTS* BY ENROLLMENT OF HIGH SCHOOLS FROM WHICH STUDENTS GRADUATED AND A.C.E. PSYCHOLOGICAL EXAMINATION QUARTILE CATEGORIES**

Enrollment of high schools from which students graduated	Mean scores on cooperative achievement tests			
	For ACE lowest quartile	For ACE second quartile	For ACE third quartile	For ACE highest quartile
	Mechanics of Expression			
Under 150	40.71	46.79	54.29	56.86
150 - 249	40.05	48.63	49.00	53.59
250 - 499	41.51	46.37	51.06	55.55
500 - 999	40.45	45.53	51.00	54.00
1000 and over	38.71	45.41	50.72	55.50
	Reading*			
Under 150	45.80	49.83	53.83	66.00
150 - 249	47.25	50.31	55.20	65.10
250 - 499	45.61	51.50	54.64	61.08
500 - 999	45.47	50.80	54.82	64.25
1000 and over	47.70	51.68	56.33	60.30
	Mathematics			
Under 150	50.05	54.51	58.47	59.57
150 - 249	49.84	55.63	55.24	61.41
250 - 499	50.36	55.17	58.20	60.65
500 - 999	49.95	53.18	56.56	60.40
1000 and over	51.39	56.15	59.00	62.43
	Science			
Under 150	51.43	54.72	54.41	60.57
150 - 249	50.70	55.51	55.00	60.71
250 - 499	50.88	55.77	58.14	59.25
500 - 999	53.68	55.53	57.31	61.60
1000 and over	51.64	56.37	59.13	59.50

---

* Scores on the reading achievement test were available for only 485 of these students.

**Agricultural Curricula, A. and M. College of Texas, 1946-48 Entering classes at Annex.

did students from the large high schools. Students in the lowest quartile category from high schools with enrollments of under 150 had a mean grade point ratio of 0.06 as compared with a ratio of -0.29 for students from high schools with enrollments of 1000 and over. Students in the second quartile category from the small schools had a mean grade point ratio of 0.83 as compared with a ratio of 0.41 for students graduating from the large high schools. Mean grade point ratios for students in the highest quartile category had little uniform relationship to enrollment of high schools from which students graduated.

Mean grade point ratios in all natural science courses taken by the end of the students' second year in college were more closely related to aptitude than to enrollment of high schools from which students graduated (Table V). However, students in the lowest aptitude quartile from the small schools had a mean grade point ratio of 0.18 which was .48 points higher than the -0.30 grade point ratio for students from high schools with enrollments of 1000 and over. In the second aptitude quartile there also appears to have been a reasonably clear negative relationship between size of schools and mean grade point ratios. For the third and for the highest aptitude quartiles there was no clear-cut relationship between enrollment of high schools and mean grade point ratios.

For the first three aptitude quartiles a negative relationship was apparent between enrollment of high schools and mean grade point ratios in all agricultural courses taken by the end of the second year. For students in the lowest quartile who graduated from high schools with enrollments of less than 150, the mean grade point ratio was 1.17, while students in the same quartile who graduated from the large high schools of 1000 or more enrollment had a grade point ratio of 0.47. Not so clear a trend was seen in the second and third quartiles, but some of the same relationship was present. In the highest quartile no relationship was evident between enrollment of high schools and mean grade point ratios in agricultural courses. There was a clearer relationship existing between aptitude scores and mean grade point ratios in agricultural courses than between sizes of high schools and mean grade point ratios.

When students' completion of the second year in college was used as a measure of academic success and compared with size of high schools, without controlling the effects of aptitude, it was seen that 62.26 percent of students from the medium sized schools with enrollments of 250 to 499, inclusively, were successful (Table VI). The next highest rate of completion was by students graduating from the large high schools with enrollments of 1000 or over. While these were the two highest rates of completion, there was no clear-cut relationship trend between high school enrollments and completion of the second year of college study.

When the effect of aptitude was controlled, there still appeared to be no consistent trend which would make it possible to use size of high school as a basis for predicting scholastic survival in college. As has already been shown, enrollments of high schools from which students graduated were related to scores on the A.C.E. Psychological Examination. Aptitude scores have also been shown to be related to academic achievement. Only in this way was there a clear-cut relationship between size of high school and academic success as measured by completion of second year in college. Students in the lowest aptitude quartile had a second year completion rate of 42.91 percent, while 74.60 percent of the students ranking in the highest quartile completed the second year.

Completion of the second year is a quantitative measure of academic achievement. By adding the qualitative measure of mean grade point ratios at the end of the second year for all students completing the second year, the two measures become more meaningful (Table VII). Because of the small number of students in the highest quartile coming from high schools with enrollment categories of under 150 and of 500 to 999, inclusively, the mean grade point ratios shown in Table VII for these two categories probably have little true meaning.

Students in the lowest aptitude quartile who graduated from high schools with enrollments of under 150 had a mean grade point ratio in all courses taken by the end of the second year of 1.17 while those from schools with enrollments of 1000 and over had a mean grade point ratio of 0.74. Except for students ranking in the lowest aptitude quartile, there was little consistent relationship apparent between size of high school and mean grade point ratios. However, within each category for size of high school, there was a direct relationship between aptitude quartile categories and mean grade point ratios in all courses taken at the end of two years.

No consistent relationship existed between the mean grade point ratio in all English courses (Table VII) for students completing the second year and size of high schools from which the students graduated. As was true for the ratios in all courses taken, there was a direct relationship within each size of high school category between mean grade point ratios in English and aptitude quartile categories. Students from schools with less than 150 enrollment and who ranked in the lowest aptitude quartile were relatively less successful in English than in all courses taken, in natural science courses or in all agricultural courses.

TABLE V

MEAN GRADE POINT RATIOS AT END OF SECOND YEAR FOR 621 STUDENTS BY ENROLLMENT OF HIGH SCHOOLS FROM WHICH STUDENTS GRADUATED, AND A.C.E. PSYCHOLOGICAL EXAMINATION QUARTILE CATEGORIES *

Enrollment of high schools from which students graduated	Mean grade point ratios at end of second year			
	For ACE lowest quartile	For ACE second quartile	For ACE third quartile	For ACE highest quartile
	In all courses taken			
Under 150	0.47	0.94	1.51	1.20
150 - 249	0.35	0.88	1.05	1.59
250 - 499	0.35	0.94	1.33	1.33
500 - 999	0.27	0.88	1.35	1.38
1000 and over	0.10	0.77	1.20	1.58
	In English courses			
Under 150	0.06	0.83	1.28	1.29
150 - 249	0.08	0.79	0.79	1.21
250 - 499	-0.02	0.41	0.75	1.25
500 - 999	0.01	0.35	1.24	0.98
1000 and over	-0.29	0.41	0.85	1.45
	In Natural Science courses			
Under 150	0.18	0.72	1.29	1.24
150 - 249	0.01	0.51	0.57	1.45
250 - 499	-0.08	0.64	0.97	1.05
500 - 999	-0.12	0.36	1.01	0.92
1000 and over	-0.30	0.53	0.87	1.40
	In Agricultural courses			
Under 150	1.17	1.57	2.02	1.83
150 - 249	0.82	1.32	1.50	1.81
250 - 499	0.97	1.45	1.62	1.50
500 - 999	0.66	1.51	1.61	1.67
1000 and over	0.47	1.23	1.53	1.89

*Agricultural Curricula, A. and M. College of Texas, 1946-48 Entering classes at Annex.

TABLE VI

DISTRIBUTION OF 352 STUDENTS* COMPLETING SECOND YEAR BY ENROLLMENT OF HIGH SCHOOLS FROM WHICH STUDENTS GRADUATED, AND BY A. C. E. PSYCHOLOGICAL EXAMINATION QUARTILE CATEGORIES**

Enrollment of high schools from which students graduated	Total No.	In ACE lowest quartile	In ACE second quartile	In ACE third quartile	In ACE highest quartile
		Number completing second year			
Under 150	83	38	27	14	4
150 - 249	70	26	20	10	14
250 - 499	99	32	24	27	16
500 - 999	31	11	7	11	2
1000 and over	69	11	26	21	11
Total	352	118	104	83	47
		Percent completing second year			
Under 150	53.90	43.68	62.79	82.35	57.14
150 - 249	52.63	40.63	57.14	58.82	82.35
250 - 499	62.26	43.24	80.00	77.14	80.00
500 - 999	51.67	50.00	41.18	68.75	40.00
1000 and over	60.00	39.29	63.41	65.63	78.57
Total	56.68	42.91	62.65	70.94	74.60

---

* Enrollment of high schools was unknown for 9 out of the total of 361 students completing the second year.

**Agricultural Curricula, A. and M. College of Texas, 1946-48 entering classes at Annex.

TABLE VII

MEAN GRADE POINT RATIOS FOR 352 STUDENTS COMPLETING SECOND YEAR BY ENROLLMENT OF HIGH SCHOOLS FROM WHICH STUDENTS GRADUATED AND BY A. C. E. PSYCHOLOGICAL EXAMINATION QUARTILE CATEGORIES*

Enrollment of high schools from which students graduated	Mean grade point ratios at end of second year			
	For ACE lowest quartile	For ACE second quartile	For ACE third quartile	For ACE highest quartile
	In all courses taken			
Under 150	1.17	1.30	1.67	0.95
150 - 249	1.08	1.37	1.41	1.68
250 - 499	1.01	1.18	1.51	1.43
500 - 999	0.87	1.63	1.57	0.99
1000 and over	0.74	1.14	1.57	1.64
	In English courses			
Under 150	0.39	0.95	1.32	0.95
150 - 249	0.68	0.99	1.07	1.29
250 - 499	0.52	0.52	0.87	1.24
500 - 999	0.40	1.21	1.25	0.32
1000 and over	0.03	0.65	1.27	1.48
	In Natural Science courses			
Under 150	0.85	0.97	1.44	1.21
150 - 249	0.68	1.02	0.90	1.44
250 - 499	0.65	0.81	1.15	1.17
500 - 999	0.61	1.48	1.41	0.54
1000 and over	0.18	0.80	1.23	1.36
	In Agricultural courses			
Under 150	1.61	1.64	2.02	1.78
150 - 249	1.42	1.77	1.67	1.91
250 - 499	1.55	1.50	1.74	1.58
500 - 999	1.22	1.97	1.78	1.35
1000 and over	1.06	1.49	1.76	1.94

*Agricultural Curricula, A. and M. College of Texas, 1946-48 entering classes at Annex.

Mean grade point ratios in all natural science courses taken by students completing the second year had little consistent relationship with size of high schools from which students graduated (Table VII), but these ratios appeared to be directly related to aptitude quartile categories within each grouping for enrollment of high schools.

For students completing the second year, little consistent relationship was found between either size of high schools or aptitude quartile categories and mean grade point ratios in all agricultural courses taken during the two years.

Summary

Three points stand out in all the tables which have been presented to show the qualitative relationship to enrollment of high schools from which students graduated. With the exception of mean grade point ratios in English, the highest grade point ratios made by students in the lowest aptitude quartile were made by students from small high schools with enrollments of under 150 students. With no exceptions, the lowest mean grade point ratios shown by any category were made by students in the lowest quartile from the large schools of 1000 or more enrollment. With few exceptions, the highest mean grade point ratios for any category were earned by students in the highest quartile who were from the large schools of 1000 and over enrollment.

BIBLIOGRAPHY

1. Ashmore, B. "High School Marks as Indicators of College Success," Journal of American Association of Collegiate Registrars, XXI (January 1946), pp.. 219-30.

2. Assum, A. L. and Levy, S. J. "Comparative Study of the Academic Ability and Achievement of Two Groups of College Students," Journal of Educational Psychology, XXXVIII (May 1947), pp. 307-310.

3. Berdie, Ralph F. "Prediction of College Achievement and Satisfaction," Journal of Applied Psychology, XXVIII (June 1944), pp. 239-245.

4. Bolenbaugh, L. and Proctor, W. M. "Relation of Subjects Taken in High School to Success in College," Journal of Educational Research, XV (1927), pp. 87-92.

5. Borow, Henry. "Current Problems in the Prediction of College Performance," Journal of the American Association of Collegiate Registrars, XXII (October 1946), pp. 14-26.

6. Bradford, H. E. Analysis of Achievement of Certain University of Nebraska Students Who Offered Vocational Agriculture as Credit for Entrance Compared with Achievements of a Similar Group Who Offered the Traditional Entrance Subjects, unpublished Ph. D. thesis, Cornell University, 1932.

7. Charles, D. C. "College Performance of Top-Quarter High School Graduates," Journal of Educational Psychology, XXXIX (February 1948), pp. 82-91.

8. Durflinger, G. W. "Prediction of College Success—A Summary of Recent Findings," Journal of American Association of Collegiate Registrars, XIX (October 1943), pp. 68-78.

9. Feder, D. D. "Factors Which Affect Achievement and Its Prediction at the College Level," Journal of American Association of Collegiate Registrars, XV (January 1940), pp. 107-18.

10. Freeman, E. M. and Johnson, P. O. "Prediction of Success in the College of Agriculture, Forestry, and Home Economics," University of Minnesota Studies in Predicting Scholastic Achievement, Part One, (Minneapolis: University of Minnesota Press, 1942), pp. 33-65.

11. Hepner, W. R. "Factors Underlying Unpredicted Scholastic Achievement of College Freshmen," Journal of Experimental Education, VII (March 1939), pp. 159-98.

12. Hurd, A. W. "Problems of the Prediction of College Success," Journal of Educational Research, XXXVIII (November 1944), pp. 217-19.

13. McNeeley, John H. "College Student Mortality Studies," Journal of American Association of Collegiate Registrars, XV (January 1940), pp. 119-24.

14. Orr, H. K. "College Achievement: Comparison of the Records Made in College by Students from Fully Accredited High Schools with Those of Students Having Equivalent Ability from Second and Third Class High Schools," Journal of Educational Research, XLII (January 1949), pp. 353-64.

15. Quarles, B. "Student Separations from College: Overview," Association of American

Colleges Bulletin, XXXV (October 1949), pp. 404-9.

16. Scott, C. M. "Background and Personal Data as Factors in the Prediction of Scholastic Success in College," Journal of Applied Psychology, XXII (February 1933), pp. 42-9.

17. Smith, F. F. "Use of Previous Record in Estimating College Success," Journal of Educational Psychology, XXXVI (March 1945), pp. 167-176.

18. University of Minnesota. Studies in Predicting Scholastic Achievement, Part I. (Minneapolis: University of Minnesota Press, 1942).

19. Veterans' Administration. "Predicting Success in Training for Agriculture," Veterans' Administration Technical Bulletin, TB 7-90 (April 28, 1948).

20. Weintraub, R. G. and Salley, R. E. "Graduation Prospects of Entering Freshmen," Journal of Educational Research, XXXIX (October 1945), pp. 116-26.

CHILDREN'S THINKING IN THE PROCESS OF SUBTRACTION[*]

E. GLENADINE GIBB[**]
Iowa State Teachers College
Cedar Falls, Iowa

Purpose and Background

THE PURPOSE of this study was to analyze children's thoughts in solving mathematical problems involving the process of subtraction. More specifically, this investigation was undertaken to study children's thinking during the "intervening interval" between the time they are asked to solve a problem and that time when they present a solution in an effort to understand the way in which children perceive applications of the process of subtraction as the degree of abstraction in the context of the number situation varies.

In the past two decades such research studies as made by McConnell (12), Thiele (21), and Swenson (20) on comparative methods of teaching found that a "meaning theory" of instruction yielded superior learning results. This theory, so named primarily from the fact that children "shall see sense" in what they learn (1) emphasizes methods of instruction which lead children to discover relationships and observe generalizations. Present theories of the teaching of mathematics as presented in *The Teaching of Arithmetic* (15) and *The Learning of Mathematics Its Theory and Practice* (14) place emphasis on making mathematics meaningful by providing children with experiences that will enable them to understand the mathematical concepts and their applications.

Judd (9) has described two ways in which a student may acquire a knowledge of what a process really means: (a) he may be given many situations in which the process is used and by a "gradual process of abstraction and generalization arrive at an understanding of the process," or (b) he may be given explicit instruction in the meaning of the process by the teacher which thus hastens his arrival at the understanding of the general idea. The latter has been used in our present instructional program for it has seemed the more economical of the two procedures. In such programs the child has been urged and perhaps too often compelled by his elders to learn rapidly and without regard for his native interests a system of numbers consisting of abstract ideas. Little is known regarding the effect of this emphasis on "products of learning" upon the most efficient utilization of a child's capacity to learn. However, if a child is to be guided effectively in the development of mathematical concepts, then it seems necessary not only to analyze the number system but also to carefully consider children's ways of thinking.

The importance of this two-fold consideration in the process of subtraction was treated by Judd (9). More recently, it has been noted by Swenson (19) in a consideration of instructional procedures in arithmetic for primary grade children and more specifically, by Van Engen (22) and Hartung and others (5) in formulating a method of instruction in this area.

From an evolution of a number system to describe the numerosity of objects, a concept of a number system has been developed apart from the context of collections of distinct objects. This abstract form has become independent of objects to be counted. Thus, if a_{s-1} is the unique solution of the equation $x + a_1 = a_s$, then a_{s-1} is the difference between a_s and a_1, or $a_{s-1} = a_s - a_1$. This process of finding a_{s-1} when a_s and a_1 are given is defined as subtraction. When numbers of a system are associated in one-to-one correspondence with distinct groups of objects, as A_1, A_s, and A_{s-1} respectively, subtraction becomes the process of finding the number associated with the difference of two groups, A_s and A_1. Or it may be said to be the process of finding the other group, which when combined with A_1 will give A_s. In identifying applications in situations familiar to children, one can show that they are all solved by the same mathematical process and that they are mathematically equivalent.

Studies in concept formation as made by Piaget (16, 17), Long and Welch (10, 11, 23, 24), Heidbreder (6, 7) indicate that meaning is, at least in part, a function of experience. The limited range of children's experiences with mathematical concepts in contrast to those of adults

---

[*] A summary of an unpublished dissertation submitted to the faculty of the Graduate School of the University of Wisconsin in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

[**] The writer wishes to express her sincere and deep appreciation to Professor John R. Mayor, under whose direction the study was made, and to Professor Chester W. Harris, under whose guidance the experiment was designed, for the assistance and encouragement given her.

suggests that children's perception of quantitative situations might be much different from that of adults. Children, therefore, may not perceive the mathematical identity in varying applications of the process of subtraction.

Thus, in the classroom if the teacher is providing experiences in subtraction as she conceives of the concept and if a child is trying to understand the situation with a different concept of the process to be used, then it may well be expected that these conflicting ideas will cause the child to become confused. In a broader sense, if mathematical concepts and their applications are developed in the classrooms from the concepts of adults, and if children bring to those same situations immature and possibly different concepts, then it may well be expected that a chaotic learning experience will result.

Statement of the Problem

This investigation was an effort to determine whether or not three problem situations to which the process of subtraction is applicable elicit different responses from children in selected second grades.

In this study thinking has been regarded as the mental activity a subject uses in recognizing relationships among properties of objects and symbols. This mental activity was observed indirectly through responses given by children in relating their manner of solving selected problems where the process of subtraction and the context of the problem were systematically varied.

Three applications to which the process of subtraction is applicable were defined as: (1) take-away subtraction situations, which presented the subject with a group from which a part was to be taken; (2) additive-subtraction situations, which presented the subject with a part of a desired group to which another part must be added that the group might be obtained; and (3) comparative-subtraction situations, which presented the subject with two groups for which the difference in the numerosity of the groups was to be determined.

The contexts, in which these applications were presented, were defined as three levels of abstraction—concrete, semi-concrete, and abstract—as commonly considered in the teaching of mathematics. (1) At the concrete level, the problem was presented in a context of small toys and other manipulative objects. (2) At the semi-concrete level, the problem was presented in a context of groups of circles and/or squares mounted on cards. (3) At the abstract level, the problem was presented in a context of language and number symbols.

In this investigation children's thinking during the "intervening interval" between the time they were asked to solve a problem and that time when they presented a solution was studied in an effort to understand the way in which children perceive applications of the process of subtraction. In addition, the degree of abstraction in the context of the number situation was varied in testing the following null hypothesis:

> For young children, certain applications of subtraction do not present situations in which children respond differently.

The principal questions which this study has attempted to answer are:

1. Do these three applications of subtraction present situations in which young children respond differently? If so, in what ways are these differences evident?
2. Do the contexts in which the applications are presented influence the children's responses to the problems? If so, in what ways are these influences evident?
3. Is one kind of context more appropriate for an application than another context? If so, in what way is one more appropriate than another?

Since this was an exploratory study, no effort was made to control experience. Rather, this investigation has attempted to analyze mental processes of a group of randomly selected second grade children in their solutions of problems in the categories as defined above.

Design of the Investigation

The Problems—For each of the nine combinations of application and abstraction, nine number pairs were used in describing quantitative situations familiar to children. These number pairs were selected at random from those subtraction facts whose minuends are less than ten and whose differences are more than one. That the nine problems for each combination might appear more than once in the investigation, each of the eighty-one problems appeared four times, thus making thirty-six problems for each type of situation or three hundred twenty-four problems in the investigation.

Small toys and other manipulative objects (blocks, marbles, pennies, nickels, jacks, balls, pencils, cards, cars, trucks, trains, balloons, whistles, paint brushes, and paints) were used in the concrete context. At the semi-concrete level, each group was pictured by dark green paper circles or squares mounted in a linear arrangement on 5 × 8 white cards. For the abstract context, problems were written about the same descriptive materials with each of the number pairs as used at the object level. These problems were typewritten (primer type) on 5 × 8 white cards.

Although the nine number pairs appeared in the same order for each child, no child had the same sequence of types of problems, and for no child did any like application appear in two contexts consecutively.

The Children—Since each child was to work a problem in each of the nine combinations of application and abstraction, thirty-six children were selected from second grades in twenty-four graded schools (two or more teachers) located in East and West Districts, Dane County, Wisconsin. One to three children were randomly selected from each school.

These children ranged in intellectual ability from intelligence quotients of 83 to that of 162, according to scores on intelligence tests given in the respective schools. The mean of intelligence quotient scores for the group was approximately 114. As indicated by the occupations of the fathers, these children came from homes representing varying levels of socio-economic status. From the information given by the teacher regarding the child's background of experience and success in arithmetic, it was noted that all thirty-six children had been taught the take-away method of subtraction. Furthermore, all had at least limited experience at the time of the interview with the applications of subtraction used in this study. The kind and amount of number experiences varied considerably.

The Interview—The writer, who conducted all interviews, and the subject were provided a room where they might work without interruption. Before the nine problems were presented, several problems were given to clarify any questions the child might have concerning the relating of the processes he used in solving the problems. All problems were presented verbally, as the respective contexts were placed before the child. Following the giving of his answer, the child was asked to tell how he arrived at his solution.

All interviews were recorded on tape and were supplemented at the time of the interview by notes taken of observed behavior of the child while he worked. It was from these recorded interviews that an analysis was made of children's responses to these nine types of problems.

Scoring Procedures and Statistical Techniques

In making an analysis of the recorded interviews, six variables, were defined in the following manner:

Process was the manner of solving the problem with reference to the level of mathematical procedure used to arrive at a solution (i.e., from simple forms of counting to the use of number facts).

Understanding was the degree to which the idea of the problem was recognized.

Equation was the level of development of the subtraction equation $a_{S-1} = a_S - a_1$, in the process of solving the problem.

Solution was the adequacy of the response given in answer to the question asked regarding the problem situation.

Time was the length of the period from the conclusion of the verbal presentation to the giving of the solution to the problem.

Verbalization was the number of words used in response to the question, "How did you find out?"

Rules were formulated for assigning a number to a level of difficulty of the property considered. A measurement for each property on each of the three hundred twenty-four problems was thus given. Three persons were asked to serve as judges in scoring fifty interviews randomly selected from the three hundred twenty-four interviews. These judges were in agreement on approximately seventy-five percent of the scores with those given by the writer and at least two judges were in agreement on about ninety-one percent of the scores. The writer re-scored this same sample and was in ninety-nine percent agreement with the original scores she had given.

The present investigation was so designed that an application of the technique of the analysis of variance might be used in testing the hypothesis. The mathematical model appropriate for this analysis was one in which both systematic and random variation are considered since applications, contexts, and variables for scoring were sources of systematic variation and variation assignable to children a source of random variation. Thus the mixed model was used in the analysis of the composite score (sum of five variables, excluding verbalization). For the analysis of each of the six variables the simple linear regression model was used. A 3 × 3 × 36 randomized block design with one observation per cell was used for each of these analyses. Where each of the five variables were regarded as a fourth dimension, a four-factor analysis was made.

Since there was only one observation per cell, there was no estimate of error from within cells. Thus, the three-factor interaction sum of squares was used in its place. Where only an overall analysis of variation contributed by the components was made, main effects and interaction effects were tested against the second-order interaction effect in the three-factor analysis, and against the third-order interaction effect in the four-factor analysis. Models used in the estimation of components of variance are found in papers by Johnson and Keeping (8), Crump (2), Daniels (3), and Satterthwaite (18). However, the particular approximate test used in this investigation for components of variance was that described by Mood (13). The sum of squares for the triple interactions, as well as for the third-order interaction was found by direct calculation as described by Edwards (4).

In all analyses the one percent level of significance was used for the rejection of the hypothesis involved.

Analysis of Results

Analyses were made of the variation (a) according to a composite score, a linear combination of five scores taken from the interviews—namely, process, understanding, equation, solution, and time; (b) according to each of the six variables considered independently; and (c) according to five variables which were no doubt correlated, considered as a group. Standard scores were used for analyses in (a) and (c) and raw scores were used for analyses in (b).

In Table I-A the analysis of variance of the behavior as represented by a composite score of the five variables is given. From this analysis one observes significant main effects in applications, contexts, and individuals (children) and in the interaction of individuals with both applications and contexts. There was no significant interaction between applications and contexts. In all cases these hypotheses are statements in the form of null hypotheses.

In Table I-B the main effects (applications, contexts, individuals) have been compared with two-factor interactions rather than with the triple interactions as was done to obtain the results in Table I-A. The rationale for the second comparison as given in Table I-B follows. It was observed in Table I-A that the null hypothesis was rejected for A × I, namely, that applications and individuals do not produce effects together which are not produced independently. Consequently, if interactions and applications were of the same order of magnitude, then one could not be assured that there was a difference between applications. In consequence, application differences are important if they are large relative to applications-individuals interaction. Similarly, context and application differences are important if they are large relative to interaction effects. From Table I-B it is noted that the application differences are most important against interaction.

A similar analysis of variance for each of the variables—process, understanding, equation, solution, time, and verbalization was made. Table II presents the analyses of variance for each of these six variables. Summarizing these analyses, differences do exist among the applications and among the contexts for each of the variables, excluding verbalization. One observes no significant interaction between applications and contexts for any variable in any analysis. However, again excluding verbalization, there is significant interaction between applications and individuals for all variables. Significant interaction between contexts and individuals was present only in the equation used and in the time taken to arrive at a solution.

In the four-way design (5 × 3 × 3 × 36) an analysis of variance was made of the variables, applications, contexts, and individuals (children). From this analysis, presented in Table III, one observes significant differences among applications, among contexts, and among individuals. No significant interaction effects are observed between either applications or variables with contexts. However, there are observed significant interaction effects between individuals with contexts, applications, and variables, respectively. Too, there is significant interaction between applications and variables. Only one significant triple interaction is observed, that of applications, contexts, and individuals. All other triple interactions were combined with the interactions of variables, individuals, contexts, and applications. Main effects and other interactions were then tested against this combined group.

The Findings—For those analyses in which the null hypotheses regarding applications and contexts were rejected, a comparison was made between the three applications and the three contexts. The findings follow:

Applications

1. Composite behavior: Take-away problems appeared easiest and comparative problems most difficult if the degree of difficulty of a problem may be said to vary inversely with the level of performance, that is, for problems less difficult a higher degree of performance may be expected while for problems more difficult a lower degree of performance is expected.

2. Understanding: The lack of understanding for comparative problems was observed in particular during the interviews. For example, when a child was presented with eight forks and five knives and the question, "How many more forks do you see than knives?" the response was commonly, "Eight forks."

3. Equation: Approximately three-fourths of the take-away problems were solved by the equation, $a_{S-1} = a_S - a_1$ while the remaining one-fourth was almost equally divided between the use of the equation of the form $a_1 + a_{S-1} = a_S$

TABLE I*

ANALYSIS OF THREE EFFECTS—COMPOSITE SCORE

Source of Variation	d.f.	A		B	
		F	Hypothesis (H_0)	F	Hypothesis (H_0)
Among Applications	2	54.47	Rejected	21.21	Rejected
Among Contexts	2	13.21	Rejected	6.07	Rejected
Among Individuals	35	7.74	Rejected	2.06	Rejected
Interactions					
A × C	4	0.39	Not rejected	0.39	Not rejected
A × I	70	2.58	Rejected	2.58	Rejected
C × I	70	2.17	Rejected	2.17	Rejected
A × C × I	140				

*For complete tables, see writer's unpublished thesis, Library, University of Wisconsin.

TABLE II*

SUMMARY OF THE ANALYSES OF THREE EFFECTS FOR EACH OF THE SIX SCORES

Source of Variation	d.f.	Process		Under-Standing		Equation		Solution		Time		Verbal-ization	
		F	H_O	F	H_O	F	H_O	F	H_O	F	H_O	F	H_O
Among Applications	2	4.62	NR**	60.00	R	51.04	R	13.80	R	8.41	R	0.41	NR
Among Contexts	2	2.07	NR	5.00	R	5.33	R	6.08	R	8.07	R	0.53	NR
Among Individuals	35	3.63	R	5.75	R	8.13	R	4.33	R	3.33	R	6.12	R
Interactions													
A × C	4	0.31	NR	1.75	NR	0.17	NR	0.51	NR	1.74	NR	0.90	NR
A × I	70	1.28	NR	4.00	R	2.83	R	1.92	R	2.41	R	1.23	NR
C × I	70	1.42	NR	1.25	NR	2.00	R	1.35	NR	2.73	R	0.92	NR
A × C × I													

* For complete tables, see writer's unpublished thesis, Library, University of Wisconsin.

**R = H_O rejected; NR = H_O not rejected; H_O = null hypothesis.

and by means of no identifiable equation. Three-fourths of the additive problems were solved by the addition equation and one-eighth of them by the subtraction equation. Approximately thirty-eight percent of the comparitive problems were solved by use of the subtraction equation, about thirty percent by the addition equation, and twenty-five percent by means such that no equation could be identified. The remainder of the problems were not attempted.

In eighty-one percent of the take-away problems the equation was inferred from what was actually done or indicated as necessary to do (e. g., taking out) in order to solve the problem. Similarly, seventy-seven percent of the additive problems and eighty-two percent of the comparative problems were solved in terms of actively thinking with the groups (e. g., putting more on, one-to-one correspondence) rather than in terms of the mathematical symbols.

4. Solution: A higher degree of rightness of solution was given for take-away problems than for the other kinds of problems. Most incorrect solutions were given to problems involving comparative situations.

5. Time: Additive problems required a slightly longer period of time than comparative problems but both took a longer period of time than did the take-away problems.

Summary—There were significant differences among applications for the understanding, equation, solution, and time scores. Highest degree of attainment was made in take-away problems and lowest level of attainment was made for comparative problems. Additive problems were found to take a longer period of time for solution and were farthest removed from the generalization of the subtraction process being distinct from addition. Thus, for seven-year-old children, types of subtraction problems do not appear mathematically or psychologically the same.

Contexts

1. Composite behavior: There was an indication of a higher level of performance for problems which were presented with stimuli in the form of squares and/or circles mounted on cards. By the same measurement there was indication that the abstract contexts were found to be the most difficult.

2. Understanding: Problems presented with picture stimuli were slightly easier to comprehend and problems presented without direct reference to objects and pictures were most difficult to comprehend.

3. Equation: For problems presented in an abstract context, the subtraction equation was used approximately twice as often as the addition equation. For ninety-five percent of the problems presented with objects, the child's account of what he did made reference to actively taking a group away or putting more on as the means of solving the problem. This was also the situation from some eighty-seven percent of the problems presented with squares and circles and seventy-three percent of problems in abstract contexts.

   Some twenty-five percent of the problems classified according to context, were solved by means not identifiable in the form of an equation. For the remaining problems presented in concrete and semi-concrete contexts, the equations $a_1 + a_{S-1} = a_S$ and $a_{S-1} = a_S - a_1$, were used approximately the same number of times in solving problems in concrete and semi-concrete contexts.

4. Solution: The use of object and picture stimuli in presenting problems in this investigation was approximately equal in effectiveness for bringing about similar degrees of rightness of solution.

5. Time: Of the total time for the solution of all problems, forty-three percent was used for problems presented with an abstract stimulus and fifty-seven percent for the object and picture contexts combined.

Summary—There were significant differences among contexts for levels of understanding, equation, solution, and time scores. Levels of performance for problems presented in semi-concrete contexts were higher than for those presented in concrete contexts. However, lowest scores were made by responses given to problems presented in abstract contexts. Therefore, contexts utilized in this study, semi-concrete contexts are less difficult than concrete contexts. Children have less difficulty solving problems if they can manipulate objects or at least think in presence of objects with which the problems are directly associated than when solving problems wholly on a verbal basis.

Individuals

There were significant differences among individuals for all six measurements. An informal analysis suggested that two factors which may be related to this difference are intelligence and experience. No attempt was made to identify the relationship or contribution of either to the responses given.

Interactions

1. Observing the composite description of each response made to each application, in no case did all applications evoke the same behavior from the same child.

2. These findings indicated that children individually pass through sequential stages, or levels of the development of the mathematical generalization of subtraction.

3. These children did not respond equally well to problems presented in varied contexts, that is, for children of this age there is no context in which problems appeared the same to all.

4. For no child were measurements similar for all variables. For the most part, each child was unique in the particular combination of scores he received for his group of nine responses. Thus, one score does not give a uniform picture of the child's behavior in problem solving situations. Not only did different children bring many varied procedures to problem solving situations, but also the same child used many varied procedures.

5. There were found no statistically significant differences in the interaction of applications and contexts. For these children the applications and contexts in which the applications were presented are independent of each other.

6. There was significant interaction between applications and individuals. Hence, this study gave no evidence of a general subtraction ability which characterizes children.

7. There was significant interaction between contexts and individuals. These children did not respond equally well to problems presented in similar contexts.

8. There was significant interaction between variables and individuals. No score gave a uniform picture of the child's response to the problems and for no group of five or more children were there similarities among the pattern of scores for the different variables.

9. There was no significant interaction between variables and contexts; that is, the scores used were not more adapted to one kind of context than to another.

10. There was significant interaction between variables and applications. The equation score indicated that comparative problems were more closely conceived as applicable of the process of subtraction than additive problems. This interpretation is in support of the first result above.

Conclusions

Within the limitations of this exploratory investigation the following conclusions were made.

The principal conclusion, the one which answered the question raised by the null hypothesis, was that relatively young children with limited number experience gave different responses to problems involving three of the applications of subtraction. They do not as a group conceive that one basic idea appears in all applications. Rather, they solved the problems in accordance with the reality of the problem situation.

Furthermore, the kinds of materials (objects, pictures, words) used in the problem stimuli suggest different responses. Children were slightly more successful in the solutions of problems when they were solved in the presence of objects. According to the measures used in this study, children more easily comprehended the problems and saw the applications in relation to the process of subtraction when problems were presented in semi-concrete contexts. Problems in either concrete or semi-concrete contexts were less difficult than problems presented wholly on a verbal basis. For materials of the type used and measurements employed in this study, children were most successful in general overall problem solving behavior with those problems presented in semi-concrete contexts.

There was no context which appeared more appropriate to one application than to another.

Suggestions for Further Research

1. All results of this investigation may be said to relate to such localities as Dane County, Wisconsin, where there exists a similar organization of school systems and the same philosophy of education. Findings for other types of localities are needed to support this study.

2. Additional evidence obtained from alterations of design would be desirable. The lack of replication in the design used in this study has been recognized.

3. A longitudinal study, not only of group trends but also of case studies would give helpful information concerning the evolving of the mathematical concept of subtraction.

4. The problem of the effectiveness of certain concrete materials in concept development seemed to be suggested by this study, raising the question as to whether or not certain kinds of materials detract from the basic concept to be developed.

5. Mathematical and psychological analyses of many other mathematical concepts are needed

if teachers of mathematics are to understand some of the problems that may confront their students in the formation of those concepts at various levels of development.

6. It is recognized that ways of studying children's behavior in problem solving situations are crude. Research in that area as well as in the measurement of such behavior would contribute to the studies of mental processes.

BIBLIOGRAPHY

1. Brownell, William A. "Psychological Considerations in the Learning and the Teaching of Arithmetic," The Teaching of Arithmetic, Tenth Yearbook of the National Council of Teachers of Mathematics (New York: Bureau of Publications, Teachers College, Columbia University, 1935), pp. 1-31.

2. Crump, S. Lee. "The Estimation of Variance Components in Analysis of Variance," Biometrics, II (February 1946), pp. 7-11.

3. Daniels, H. E. "The Estimation of Components of Variance," Journal of the Royal Statistical Society (supplement), VI (1939), pp. 186-197.

4. Edwards, Allen L. Experimental Design in Psychological Research (New York: Rinehart and Co., 1950), pp. 174-302.

5. Hartung, Maurice and others. Numbers in Action, Teacher's Edition (Chicago: Scott, Foresman, and Co., 1951), pp. 147-150.

6. Heidbreder, Edna F. "The Attainment of Concepts, II; The Problem," Journal of General Psychology, XXXV (July 1946), pp. 191-223.

7. _______. "Problem-Solving in Children and Adults," Pedagogical Seminary and Journal of Genetic Psychology, XXXV (December 1928), pp. 522-545.

8. Johnson, Leroy P. V. and Ernest S. Keeping, "Composite Mean Squares and Their Degrees of Freedom," Applied Statistics, I (November 1953), pp. 202-205.

9. Judd, Charles H. Psychological Analysis of the Fundamentals of Arithemetic (Chicago: University of Chicago, 1927), 121 pp.

10. Long, Lewis and Livingston Welch. "Influence of Levels of Abstraction on Reasoning Ability," Journal of Psychology, XIII (January 1942), pp. 41-59.

11. _______. "Reasoning Ability in Young Children," Journal of Psychology, XII (July 1941), pp. 21-44.

12. McConnell, T. R. "Discovery versus Authoritative Identification in the Learning of Children," University of Iowa Studies in Education, IX (September 1934), 142 pp.

13. Mood, Alexander M. Introduction to the Theory of Statistics (New York: McGraw-Hill Book Co., 1950), pp. 289-364.

14. National Council of Teachers of Mathematics. "The Learning of Mathematics, Its Theory and Practice," Twenty-First Yearbook (Washington, D.C.: National Council of Teachers of Mathematics, 1953), pp. 69-98.

15. National Society for the Study of Education. "The Teaching of Arithmetic," Fiftieth Yearbook of the National Society for the Study of Education (Chicago: University of Chicago Press, 1951), pp. 53-75.

16. Piaget, Jean. Judgment and Reasoning in the Child (New York: Harcourt, Brace and Co., 1928), 260 pp.

17. _______. La Géométrie Spontanee de l'Enfant (Paris: University Press of France, 1948), 514 pp.

18. Satterthwaite, F. E. "An Approximate Distribution of Estimates of Variance Components," Biometrics, II (December 1946), pp. 110-114.

19. Swenson, Esther J. "Arithmetic for Pre-School and Primary Grade Children," The Teaching of Arithmetic, Fiftieth Yearbook of the National Society for the Study of Education (Chicago: University of Chicago Press, 1951), pp. 53-75.

20. _______. "Organization and Generalization as Factors in Learning, Transfer, and Retroactive Inhibition," Learning Theory in School Situations (Minneapolis: University of Minnesota Press, 1949), pp. 9-39.

21. Thiele, C. L. The Contributions of Generalization to the Learning of Addition Facts, Contributions to Education, No. 763 (New

York: Bureau of Publications, Teachers College, Columbia University, 1938), 84 pp.

22. Van Engen, Henry. "Teaching the Mathematical Skills in the Elementary Schools," Progressive Education, XXIX (October 1951), pp. 17-20.

23. Welch, Livingston and Lewis Long. "A Further Investigation of the Higher Structural Phases of Concept Formation," Journal of Psychology, X (October, 1940), pp. 211-220.

24. ________. "The Higher Structural Phases of Concept Formation of Children," Journal of Psychology, IX (1940), pp. 59-95.

AN EXPERIMENTAL STUDY OF CASE-DISCUSSION LEARNING

LEONARD A. OSTLUND
Kent State University
Kent, Ohio

I. The Problem

THE PURPOSE of this study was to evaluate certain goals of the case-discussion method of teaching. The content of this type of course consists of problems in interpersonal relations which occur in the family, college and industry. Students are provided with case-books and readings and encouraged to participate in classroom discussion. The teaching technique is largely non-directive, and the teacher attempts to set a permissive atmosphere in which each student will feel free to participate.

The stimulus for this type of course was a seminar conducted by Wallace B. Donham, Dean of the Harvard Business School. The writer became intrigued by the Human Relations Course while teaching it at Colgate University with Dr. F. K. Berrien, Chairman of Human Relations and author of Comments and Cases on Human Relations (5). As a result of a seminar with Dean Donham the writer became further interested.

II. History of the Problem

A survey of the literature reveals that many educators and psychologists have become interested in research concerning case-discussion teaching. In general, research has been in the area of the lecture-discussion method of teaching vs. case discussion teaching. This controversy has also been couched in terms of curriculum-centered vs. student-centered theory. The writer has contributed two research studies of case-discussion classes which have raised problems that have been incorporated in the present study (8), (9).

A recent summary by Wispe has integrated much material concerning this controversy (12). In general, the traditional lecture method of teaching has proved useful in mastery of content, but apparently non-lecture methods do equally well. Concerning favorable changes in attitudes and interpersonal skills, non-lecture methods often prove superior to traditional lecture methods. The non-lecture method effects changes of attitudes in the direction of a positive valence for subject matter and also tends to improve interpersonal skills.

III. The Experimental Design

The experimental and control groups were rigidly matched according to the following characteristics: friendship, grade point average, class standing, and previous experience in case-discussion courses. The S.'s were initially contacted by means of a questionnaire circulated in the classroom with the consent of the teachers. On the basis of this information the groups were matched and each individual was telephoned by the writer. No coercion was used. The S.'s met in groups of six at the beginning of the term and again three months later. After the final meeting a general session of all participants was held, at which time the students were acquainted with further details of the project and there was a drawing for awards. The cooperation of the students and faculty throughout the study was exemplary.

The experimental S.'s were drawn from case-discussion courses, whereas the control S.'s were drawn from non-case-discussion courses. The experimental S.'s, during the period between tests, were having experiences in the case-discussion method of instruction, whereas the control group was not. Therefore, it would be appropriate to attribute changes which occurred in the experimental group to case-discussion learning, provided these changes did not occur in the non-case-discussion learning group. Thus the control group functions as a constant frame of reference against which changes in the experimental group may be evaluated. Since the control group has not had the experience of this special educational training, we cannot expect any changes in their behavior.

The control group thus provides a base-line as it were, against which we may evaluate the progress of the experimental group. We can determine whether they have been able to transfer their case-discussion learning to this new situation. This experimentally created situation was purposely unstructured so that each S. was provided with an equal opportunity to demonstrate any general increase in maturity of approach to the discussion and solution of problems in interpersonal relationships, which had occurred over this three-

---

* This paper was adapted from a thesis written for the Ph.D. degree at the University of Kansas, Lawrence. The thesis committee comprised of Professors Fritz Heider, Marston McCluggage, Hilden Gibson, and the Chairman, Anthony Smith gave generous assistance and my wife expedited every phase of the study.

month interval. The experimental design used in the present experiment has been reproduced in Table I.

The number of S. 's per group was held constant at six for each meeting. There were six experimental (case-discussion method) groups which met for two sessions, one at the beginning and one at the end of the semester. Six control (non-case-discussion method) groups did likewise. The "after" sessions occurred three months later than the "before" sessions. The same S. 's came to the "before" and the "after" sessions and the personnel for each group was kept the same for both of these meetings. With few exceptions which were occasioned by drop-outs and refusals, almost all of the seventy-two S. 's each came to two, one-hour meetings.

The Experimental Groups—The experimental groups were chosen from four courses taught by the case-discussion method: Cases in Human Relations, Personnel Management, Cases in Marriage and the Family, and Business Administration Practices. The cases used in these courses all involve problems of interpersonal relations. While the teachers differ in regard to age, experience, and classroom technique, these courses spring from a common source—the seminars given at the Harvard Business School by Dean Wallace B. Donham—and adhere to a common set of underlying principles and methodology.

There were six groups of experimental S. 's and each group met at the beginning and at the end of a semester. Each group consisted of six S. 's who were drawn from different sections of the same course or from different case-discussion method courses, depending upon availability of the S. 's or their suitability in terms of the matching criteria.

The Control Groups—The control group S.'s were drawn from classes whose students were registered in five different schools of the University in order to gain an adequate sample of the "average non-case-discussion course student." These schools included the College of Liberal Arts, the School of Engineering, the School of Journalism, the School of Education, and the School of Business.

It was felt that balancing the personnel within a control group of six non-case-discussion students by selecting them from different schools would prove more representative of "the average non-case-discussion student." Therefore, each control group was arranged so that no more than two members, in each group of six S. 's, were from the same school. Moreover, these two individuals were chosen from different courses within the school. This arrangement was possible because there was a wide range of non-case-discussion courses in each of these five schools.

There were six groups of control S. 's, and each group of the same S. 's met at the beginning and at the end of a semester. Each group consisted of six S. 's who were drawn from different sections of the same course or from different non-case-discussion courses.

The Number of S. 's Per Group—The choice of six S. 's per group was not an arbitrary one, but was governed by important considerations. Practical considerations involved the fact that the size of the group could not be too large, since not enough case-discussion courses were available from which S. 's could be drawn so that the S.'s would come from a different section or course. Then too, the group could not be too large so that E., working alone, might not be able to observe all the interactions. However, if the groups were too small, the results would not be amenable to statistical treatment. Groups of six S. 's have been used by Bales (2), Stephan (10), and French (6), and have proved satisfactory.

After the experimental procedure had been formulated, E. tried it out in a pilot study in order to become familiar with the routine and to discover whether the methodology would prove adequate. The pilot study proved useful in that it demonstrated the adequacy of some procedures and led to some fruitful modifications.

IV. The Experimental Procedure

Barker and Wright have called attention to the coerciveness of the behavior setting—the place in which behavior takes place (3:17). With this principle in mind, E. felt that discussion would be enhanced if it took place in an appropriate physical setting. Therefore, all meetings were held in a room which was regularly used for seminars and conferences.

E. sat at one end of the table, the six S. 's sat at the other end, three facing three. This arrangement has been suggested by Bass and Klubeck (4) and used by many investigators. S. 's were given identification numbers from one to six. These were placed on the table in front of them. There was no attempt to conceal the microphone and recording equipment.* It was explained that recording was necessary in order to insure accurate data. E. kept a record of all remarks by noting the S. 's number and an identifying word. Playback of the recordings permitted each conversation to be traced to its source and insured accurate transcription.

---

* An Ekotape Tape Recorder and a Gray Audograph Disc recorder were used and proved satisfactory.

TABLE I

THE EXPERIMENTAL DESIGN

	Control Groups	Experimental Groups	Total Meetings
Before	6	6	12
After	6	6	12
Total Meetings	12	12	24

TABLE II

TABULATION OF THE TRANSCRIPT RATINGS

	1 Insight		3 Continuity		4 Atmosphere		5 Task-Centered		6 Group Centered		7 Individual Centered		8 Successful Communication	
	C.	E.	C.	E.	C.	E.	C.	E.	C.	E.	C.	E.	C.	E.
1. $\overline{X}$ – "before"	10.18	8.18	9.32	6.83	10.90	8.69	11.23	9.81	7.32	5.83	5.75	6.90	9.99	9.22
2. $\overline{X}$ – "after"	7.69	6.61	8.04	6.94	9.50	9.37	11.26	10.58	6.35	6.72	5.58	4.89	8.33	8.74
3. $\overline{X}$ – gain score	-2.49	-1.57	-1.28	.11	-1.40	.68	.03	.77	-.97	.89	-.17	-2.01	-1.66	-.48
4. $\overline{X}$ – gain difference, E. vs. C.*		.92		1.39		2.08		.74		1.86		-1.84		1.18
5. "t" test scores**		.70 (P. < 26) not significant		.63 (P. < 28) not significant		2.53 (P. < 03) significant		.35 (P. < 37) not significant		.92 (P. < 20) not significant		1.16 (P. < 15) not significant		.47 (P. < 33) not significant

---

* It will be noted that in every case, the E. group has excelled the C. group.

**All of the "t" tests are one-tail tests. A "t" of 2.02 is required for significance at the .05 level.

The leaderless group discussion technique was used. The S.'s were given the problem to read and then discussed it. Ansbacher (1), Bass (4), and others have found this a useful tool for assessing social skills of group discussion. The S.'s were strangers so that a "pure" informal group structure might emerge.

There were several reasons for using the leaderless group discussion technique. In the first place, participation by E. would constitute a familiar experience for the case-discussion-experienced S.'s, and they might interact more freely than would the non-case-discussion S.'s. However, the non-case-discussion S.'s. would not be accustomed to the situation. Moreover, any participation in discussion by E. would have necessitated a standardized procedure, the same remarks would have had to be made in the same sequence for all groups. This would have been impossible since E. was busy taking notes.

As each S. appeared for a session he was given a copy of a magazine to read while waiting. When the group had gathered E. collected the magazines, seated the S.'s and proceeded with the experiment. E. mentioned that the session was being tape recorded to insure accuracy, but that all information would be kept strictly confidential. E. placed numbers from one to six in front of each participant, explaining that S.'s could be referred to by number should anyone so desire. A "warm-up" case was distributed and discussed for five minutes. At the conclusion of this discussion the first experimental case was distributed, read, and discussed for fifteen minutes.* E. then said, "Write what you consider the best solution." When this had been done the S's completed rating scales concerning leadership in this particular situation, Communication Data Forms, and stated their opinion of the problem.

None of the twenty-four meetings was interrupted or interfered with in any way, and rapport was excellent. Once begun, all groups proceeded according to the standard methodology. The first group met in September, the last in May of the same year. The usual elapsed time for a group meeting was approximately fifty minutes, distributed as follows:

Introductory remarks	5 minutes
Problem I	5 minutes
Problem II	15 minutes
Written solution	10 minutes
Forms and rating scales	10 minutes
Concluding remarks	5 minutes

V. The Evaluation Instruments

A. The Communication Data Form

On the Communication Data Form, the S. was asked to state those with whom he specifically tried to communicate verbally and whether or not he believed that he had been successful. He was also asked to state those whom he perceived as specifically trying to communicate with him verbally, and whether or not these attempts had been successful.

B. The Leadership Rating Scale

At the conclusion of the group discussion, each S. rated every member of the group, including himself, on a Leadership Rating Scale. The S. was instructed to rate the members according to leadership in this particular situation. This made it possible to compare the perception of each member with the group's perception.

C. The Judgment of the Written Solutions

The written solutions were typewritten on separate pages, and presented in random order to eleven judges.** The judges were kept purposely unaware of the experiment. "The Badminton Problem" was used at all "before" sessions and "The Parasol Problem was used at all "after" sessions. Since the judgment criteria for both problems differed only slightly, that for "The Parasol Problem" has been omitted.

1. "Centering" Category

Rationale

"The Badminton Problem" was analyzed according to Wertheimer's concept of "centering". Wertheimer's principle is that adequate problem-solving demands centering on the total situation. Wertheimer explains centering as follows:

> "Centering—the way one views the parts, the items in a situation, their meaning and rôle as determined in regard to a core, or radix—is a most powerful factor in thinking."
>
> "As to the concept of centering, it seems to be tacitly recognized that

---

* The case used at all "before" sessions was "Two Boys Play at Badminton" (11, p.124-126). At all "after" sessions the case used was "The Parasol Problem" (7,p. 301-326).

**The writer is indebted to the following individuals who served as judges: Graduate students: Dorothy Codkind, Kelly Wiltse, Charles Buchanan, Barry Crites, Charles Deeble, Granville Fairchild, Orville Harsch, Wendell Walker. Professors: Mary Kinnane, H. K. L'Ecuyer, Martin Scheerer.

adequate centering, with its consequences for objectivity and justice is of the utmost importance'' (11:136).

Criteria

a. ''A'' (older boy) centered
b. ''B'' (younger boy) centered
c. Game or ''both'' centered
d. ''Outside of the field'' solution

2. ''Inadequate Solutions'' Category

Rationale

In his discussion of classroom research, Wertheimer pointed out that frequently the students handled the problem on a superficial level, such as labeling behavior in stereotypes, pronouncing moral judgments, and the like. The writer considered that such behaviors represented pitfalls or errors in problem-solving and this category was set up in order to detect such errors.

Criteria

a. Pigeon-holing
b. Moralizing
c. Ignoring restrictions
d. Using a restricted frame of reference
e. ''Reinterpreting'' facts

3. ''Means'' Category

Rationale

By employing this category, E. endeavored to determine the underlying means with which the S. was attempting to cope with the situation. This index of behavior is important because it tells us much about the S.'s basic viewpoint toward interpersonal relations.

Wertheimer was concerned about the fundamental approach of individuals to problems in interpersonal relations. He described the typical modes of approach to ''The Badminton Problem'' and stated that the following were the usual types of suggestions offered by his S.'s: Offering candy, playing another game, scolding, ignoring, coercing. (11:128).

It will be noted that the categories which follow have been designed to encompass many of these suggestions which Wertheimer has reported.

Criteria

a. Bargaining
b. Bribing
c. Flattering
d. Shaming
e. Mutual means
f. Withdrawal or quitting

This was the last of the three categories designed for recording the solutions to "The Badminton Problem."

Reliability of the Judges

In the present study, each written solution was judged independently by two different judges. Reliability was approached according to agreement per each category by both judges. Concerning ''The Badminton Problem'' judges' agreement was as follows: ''Centering'' Criteria—Agreement 58 percent, ''Inadequate Solutions'' Criteria—Agreement 65 percent, "Means" Criteria—Agreement 55 percent. Concerning ''The Parasol Problem'', ''Concrete Action Proposals'' Criteria—Agreement 94 percent, ''Centering'' Criteria—Agreement 66 percent, ''Inadequate Solutions" Criteria—Agreement 57 percent, ''Means'' Criteria—Agreement 74 percent. While a higher degree of agreement would have been desirable, it should be pointed out that all agreement criteria were quite rigorous, and the data was very difficult to judge.

D. The Judgment of the Group Discussion Transcripts

It will be remembered that each of the twenty-four experimental discussion sessions was tape recorded and transcribed. It is true that a transcript is a reproduction of actual behavior and subject to limitations inherent in any facsimile; however, these transcripts were made under rigorous experimental conditions, and it was not feasible to have observers present noting the interactions.

Rationale

The transcript judgment design involved counterbalancing in order to account for three important variables which might bias judgment. They were:

(1) The type of group, whether experimental or control.
(2) The time of discussion, whether ''before'' or ''after.''
(3) The problem discussed, whether ''The Badminton Problem'' or ''The Parasol Problem.''

As a preliminary measure, the judges were kept unaware of these three aspects. Furthermore, each judge judged four transcripts in a prescribed order. The "ABBA" sequence was used, in order to counterbalance the variables just mentioned.

As an additional precaution, each judge was told his set of four transcripts might include four, three, or two of one kind of group (experimental or control). Actually, each judge received the four transcripts from one matched pair, namely: Experimental Group No. 1, "before" and "after" sessions, and Control Group No.1, "before" and "after" sessions. It was believed that this combination of transcripts would result in the complete set of one matched group being judged in the same frame of reference.

E. was very fortunate concerning the important problem of gaining the assistance of competent judges. Twelve members of the University of Kansas Staff voluntarily gave generously of their time so that this task could be accomplished. Six of the judges were psychologists, six were teachers experienced in the case-discussion method of teaching. *

Criteria

The judges rated each of the four discussion transcripts according to criteria generally accepted as indicative of group functioning. Ratings were made on a continuum according to whether the group was "high" or "low" on each variable. The seven variables judged were as follows: Insight, continuity, atmosphere, task-centeredness, group-centeredness, individual-centeredness, and success of communication.

Reliability of the Transcript Judgments

With regard to agreement concerning type of group, neither the case-discussion teachers nor the psychologists who acted as judges proved superior in being able to judge whether or not the transcript was from a case-discussion or a non-case-discussion group. Moreover, all of the judges were correct in their judgment concerning the "type of group" criteria more than half of the time. This may be interpreted as indicating that the different types of groups did not have outstanding distinguishing characteristics and that neither type of judge proved clearly superior.

Concerning the seven continua ratings, agreement was analyzed by computing a Spearman Rank Difference Correlation Coefficient between the ranks assigned by the two judges to the seven variables. The results were an average Spearman Rank Difference Correlation Coefficient of .36. Thus reliability was low, and as was the case concerning the written solutions, either this type of material is difficult to judge or it was judged in different frames of reference

VI. Analysis and Interpretation

In this section the major findings from the present study will be analyzed and integration into a meaningful pattern. It should be pointed out that all the conditions and assumptions inherent in the use of the "t" test could not be met. Because random sampling was not feasible, alternative procedures were used. The universe of potential experimental subjects was relatively small so subjects were chosen in accordance with certain rigorous criteria, rather than according to personality or convenience. Control subjects who met the same criteria were chosen from a large pool in a random manner. Furthermore, restrictions of time and scope limited the study to one college and a duration of one year. However, the "t" test was used because it seemed the most appropriate. The limitations which have been noted functioned to render the conclusions tentative.

The Communication Data Analysis did not support our hypothesis concerning the predicted superiority of the case-discussion group, for the results indicated no significant differences. It may be that these skills are difficult to master, and that three months is not enough time for significant changes to occur. It is possible that this measure of communication was not adequate.

The first discovery of differences between experimental and control groups came from analysis of the "leadership" rating scale. It was found that there was no significant difference between the groups pertaining to the perception of the individual S. vs. the groups' perception of the "leadership" hierarchy. However, further analysis indicated that this may have been due to the fact that the range of "leadership" ratings for the experimental groups was much more narrow than that of the control groups. It was determined that this difference was almost statis-

---

* The writer is indebted to the following individuals who served as judges: Professors Louise Barker, Lawrence Bee, Carroll Clark, Hilden Gibson, Fritz Heider, Marston McCluggage, Lee Meyerson, Edward G. Nelson, William Thompson, Charles Warriner, Edward Wike, M. Erik Wright.

tically significant. (P. < .09). This was interpreted as suggesting that the experimental groups had excelled the control groups in that the experimental groups tended to share leadership functions. This outcome was consistent with other predictions that were made.

An analysis of the written solutions to "The Parasol Problem" provided four trends in the predicted direction, one contrary trend, and one inconclusive finding. None of these was statistically significant. The first trend involved attempts to keep the worker on the same work team. This may be interpreted as evincing a desire to contain the problem—to solve it with the minimum shifting or disturbance of personnel and interpersonal relationships. The second trend showed that the experimental group produced more solutions that attempted to use the worker productively. This may be interpreted as demonstrating a desire to use the potential of the individual to bring satisfaction to the worker and to the company.

This inference is supported by the third trend, in which the experimental group excelled the control group by centering the problem in the most adequate manner. This may be interpreted as evidence that the experimental S.'s tended to take into account the total situation, rather than approach the problem from a restricted viewpoint. It will be remembered that Wertheimer emphasized the importance of centering by stating that adequate problem-solving demanded adequate centering.

The fourth trend in which the experimental group demonstrated superiority was the finding that the experimental group produced fewer solutions that violated conditions which were stated in the problem. This was interpreted as meaning that the experimental S.'s were more willing to accept the limitations of the situation. It is generally agreed that this is a more realistic approach to problem-solving, than an approach which distorts the problem situation so that it may fit a particular solution. The latter approach may be considered an evasion of the task.

This last finding permits the resolution of alternate interpretations which were possible concerning the S.'s opinions of the problems. Our analysis indicated that the experimental S.'s voiced the criticism that relevant information was lacking more often than did the control S.'s. The alternative interpretation was proposed that either they were unwilling to accept the limitations of the situation, which was considered an unfruitful viewpoint, or that they keenly sensed that important information was lacking, which was considered a reasonable criticism. The fourth trend supports the latter interpretation. By their behavior, in the task of writing solutions, the experimental group demonstrated that they were willing to accept the limitations of the situation, and their criticisms may be interpreted as meaning that they were more keenly aware of the lack of certain information than were the control group.

In one respect the groups showed no differences. Both achieved identical scores concerning inadequacies which were detected by the judges in their solutions. The one contrary trend stands out clearly. The experimental groups produced a greater number of solutions which involved retirement. This was considered a drastic method of solving the problem since other alternatives were possible. It should be stated that no S. proposed retirement as the only solution, retirement was presented as an alternative. Therefore, it may have been considered as a last resort if other proposals failed. In this context, retirement is preferable to discharge, for retirement would benefit the worker by providing an income and would benefit the company by freeing them to hire a capable replacement. However, it would seem that the psychological repercussions—the impact on the worker and the effect of this action on company-worker relationships, — might outweigh these benefits.

The conclusion drawn from analysis of the written solutions was that the margins of superiority of the experimental group were not large. This may be considered a reflection of the difficulty of the problem. Perhaps we should not expect an individual to solve adequately a problem which six individuals acting as a group have proved incapable of solving after fifteen minutes of discussion. This aspect of the study may be summarized by stating that the experimental group excelled the control group on four measures, was excelled by the control group on one measure, and paralleled their performance on another one.

In the matter of group discussion transcripts, there were no differences between groups on the variable that pertained to the emergence of a solution from the group discussion, and the adequacy of such a solution—not one group achieved a solution, all ran out of time at the "after" session. On each of the remaining variables, the "t" tests and inspection of the relative mean gain differences indicated a consistent trend. The most striking finding of the present study was that the experimental group excelled the control group on each of the following seven variables: insight, continuity, atmosphere, task-centeredness, group-centeredness, individual-centeredness, and success of communication (Table II). Since it is generally agreed that these variables are crucial factors of group discussion processes, this performance is important. The score for the variable, "atmosphere" was statistically significant (P. < .03).

It should be mentioned that one of these variables, "insight" is more directly related to problem-solving than are the others. The margin of excellence on "insight" was low, which seems consistent with the other low findings related to problem-solving. The fact that the experimental group functioned more adequately in group discussion processes indicates that they have made more progress in this direction than in the direction of group problem-solving.

The transcript judgment analysis suggests that the overall performance of the experimental group surpasses that of the control group, that the former group has made more progress in group discussion skills. This is important, for it is generally agreed that these skills are prerequisite to group problem-solving skills.

VII. Recommendations for Curriculum Change

The fruitfulness of a study requires that it yields implications which extend beyond the experimental situation. The writer trusts that the study will prove heuristic, and believes that the results warrant the suggestion of certain practical applications and further experimental possibilities.

In the matter of curriculum improvement, the data tend to give some support to the contention of other studies that case-discussion courses provide valuable experiences that are not usually gained in the typical lecture method course. Concerning case-discussion courses of the Human Relations type, it seems that it might prove profitable to extend the training period, alter the type of training, or effect some combination of these possibilities.

It often happens that training in case-discussion classes cannot afford all members equal opportunities to learn group interaction process and group problem-solving skills because large enrollments curtail individual participation. Our results indicate that students need to become accustomed to vis-à-vis relationships apart from the classroom situation, so that a transfer of learning will be effected to areas more within the scope of daily interpersonal relationships in college, in work, and in social situations. Unless this happens, students may be content to permit the teacher to assume responsibility for furthering group interaction processes. Our results seem to indicate that a promising beginning has been made, but there is no room for complacency, further progress is essential.

Perhaps more participation, and experience in small groups could be managed by increasing the training period and instituting changes in methodology. Under the guidance of the teacher, a permissive atmosphere for classroom discussion could be established at the beginning of the course. Later, the classroom group could be encouraged to institute its own observer-feedback-evaluation procedure and these functions could be shared by the members.

Concomitantly, students could be provided with experience in functioning as members of small groups, so that each individual would have a greater opportunity to function as a responsible group member. The institution of small, leaderless groups would tend to maximize individual responsibility. It is hoped that these suggestions do not appear too unrealistic, at least some combination of these possibilities might prove feasible. Finally, if some of these suggestions were put into practice, further experimentation would become imperative in order to test the effectiveness of such procedures.

VIII. Conclusion

The gains made by the experimental group have been modest and they have been, with few exceptions, in the predicted direction. Statistically only one of these gains was significant but this must be considered in the context of the situation in which they were accomplished.

The experimental situation represented a novel experience that was a wide departure from classroom routine as practiced in case-discussion method courses from which the experimental S.'s had been drawn. Moreover, there is evidence that "The Parasol Problem" which was the crucial test, was a very complex problem. Finally, one must take into account the brevity of the training period—only three hours per week for three months—and the fact that this type of training was diametrically opposed to traditional methods.

When these conditions are taken into consideration, the writer believes it would be unrealistic to expect sweeping advances. In the writer's opinion, an appreciation of these aspects leads one to conclude that these modest gains represent an important achievement which may be attributed to the case-discussion method of teaching.

REFERENCES

1. Ansbacher, H. L., "The History of the Leaderless Group Discussion Technique," Psychological Bulletin, Vol. XLVIII, No. 5, September, 1951, pp. 383-391.
2. Bales, Robert F., et al, "Channels of Communication in Small Groups," American Sociological Review, Vol. XVI, No. 4, August, 1951, pp. 461-468.
3. Barker, Roger G. and Wright, Herbert F., Methods in Psychological Ecology, Lawrence: University of Kansas Press, 1950.

4. Bass, Bernard M. and Klubeck, Stanley, ''Effects of Seating Arrangement on Leaderless Group Discussions,'' Journal of Abnormal and Social Psychology, Vol. XLVII No. 3, July, 1952, pp. 724-727.
5. Berrien, F. K., Comments and Cases on Human Relations, New York: Harper and Bros., 1951.
6. French, John R. P., Jr., ''Organized and Unorganized Groups under Fear and Frustration,'' University of Iowa Studies in Child Welfare, Vol. XX, 1944, pp. 231-308.
7. Maier, Norman R. F., Principles of Human Relations, New York: John Wiley and Sons, Inc., 1952.
8. Ostlund, Leonard A., "A Sociometric Matrix,'' Proceedings of The Oklahoma Academy of Science, Vol. XXXIV, 1953.
9. Ostlund, Leonard A., ''Group Integration in a Case-Discussion Course,'' Journal of Educational Psychology, Vol. XLIV, No. 8, December, 1953, pp. 463-474.
10. Stephan, Frederick, "The Relative Rate of Communication Between Members of Small Groups," American Sociological Review, Vol. XVII, No. 4, August, 1952, pp. 482-486.
11. Wertheimer, Max, Productive Thinking, New York: Harper and Bros., 1945.
12. Wispe, Lauren G., ''Evaluating Section Teaching Methods in the Introductory Course,'' Journal of Educational Research, Vol. XLV, No. 3, November, 1951, pp. 161-186.

DO PROBATIONARY COLLEGE FRESHMEN BENEFIT FROM COMPULSORY STUDY SKILLS AND READING TRAINING?

WALTER S. BLAKE, JR.
University of Maryland

WHEN THE facilities of the University of Maryland were extended to all prospective students below certification average in 1947, university policy dictated that students admitted without the certification average be put on probationary status and given study and reading skills training. As a result, the persistent and important problem of how to assist college freshmen make a good adjustment in the college melieu has received a great deal of attention at Maryland University[1].

The problems are multi-causal and therefore complicated, but not essentially different than those faced by college freshmen before World War II[2]. At Maryland University, the problems have been made more acute by the ever-increasing number of applicants for admission, plus the mediocre quality of training being given in too many of the primary and secondary schools in the states from which the University draws its students. Two types of students have been put on probation since the program's beginning: (1) applicants for colleges demanding a higher level of academic achievement than high school records indicate will be attainable to the applicants; and (2) students who are dropped from one college for scholastic deficiency and who wish to transfer to another degree-granting college upon being readmitted by the Committee on Admissions, Guidance, and Adjustment. If a student thus admitted fails to earn a 2.0 (C) average over two semesters he remains on probation until he does earn the 2.0 average.

The Maryland program is unusual in that few voluntary enrollees are accepted. Present facilities do not permit further expansion, although many requests for admission are received from non-probationers and/or their advisors. Upwards of 800 students are enrolled, representing every college on campus. Lectures and laboratory sessions comprise the first semester study skills course (College Aims I); combined lecture and laboratory experiences are used in the reading improvement course given the second semester (College Aims II). Tutoring in English, as well as Mathematics is offered the students and is available upon request of students. Remedial reading instruction and counseling interviews complete the program services, all of which are offered primarily to help students learn to work up to their capacities and effect a good social adjustment in college and all their lives.

In 1952-1953 the program was evaluated in an effort to determine, in part, whether probationary college freshmen do benefit from the compulsory study skills and reading-improvement training. Two main areas were investigated; (1) the academic performance of students during and following the training given; and (2) the students' opinions of the program. The academic performance of the students was measured in several ways. The sociology grades of 128 probationers after training in study skills averaged 70.47 for one semester as compared to 75.60 for 122 members of a control group. Reading tests revealed that the average reading speed of 139 probationers was 247.3 words a minute before training began, and had improved to 406.0 at the conclusion of the reading course, with comprehension remaining practically unchanged.

One follow-up study[3] was made to determine the degree of academic success experienced by 128 probationers and 122 control-group students, over a four and one-half year period. 22.6% of the probationers graduated in the 4-1/2 year period as compared to 22.9% of the control group. Also, of the original 128 probationers who attained the 2.0 (C) average necessary to enable them to get off probation, 47.3% went on to graduate (14.5% of this group was still in school at this writing.)

A follow-up study of a second group of probationers is summed up in Table I. This table clearly indicates that a large number of lower-ranking members of the group do succeed. It would seem that the academic success of the probationers was considerable, at least partly due to the training provided in the study and reading skills program.

The anonymous student evaluation of the program was particularly enlightening. Students would rather be admitted on a non-probationary

---

[1] This article is based upon a doctoral study completed in 1953 by the author which is one of the bases for the textbook, College Orientation (2nd printing), coauthored with George Weigand, published by Prentice-Hall, Inc., 1955.

[2] Blaesser, Willard W., and Everett H. Hopkins, Counseling College Students During the Defense Period. Washington, D. C.: Office of Education, Federal Security Agency, 1952.

[3] Mrs. Martha Maxwell, assistant to the director of the program, conducted this study.

TABLE I

A. C. E. * TEST RANKINGS AND FIVE YEAR PROGRESS OF 100 PROBATIONARY STUDENTS

Percentile Rank on A. C. E.	Graduated		Withdrew		Dropped		Still in School		Total	
	N.	%	N.	%	N.	%	N.	%	N.	%
80	2	28.6	5	71.4	0	0.0	0	0.0	7	100.0
60	2	40.0	1	20.0	2	40.0	0	0.0	5	100.0
40	6	27.3	3	13.6	10	45.4	3	13.6	22	100.0
20	8	33.3	4	16.7	10	41.7	2	8.3	24	100.0
0	3	7.1	15	35.7	21	50.0	3	7.1	42	100.0
Total	21	21.0	28	28.0	43	43.0	8	8.0	100	100.0

*American Council on Education Psychological Examination rather than probationary basis when they apply for admission to a college or university, of course; but, once put on probation and required to complete a study skills and reading improvement program, do students then recognize that the experience has helped them? A systematic and entirely anonymous student evaluation was undertaken at the University of Maryland in an attempt to answer this question. Over a period of three semesters (1951-1953) 599 participants in the program were asked to evaluate their experiences anonymously: 72 were enrolled in a lecture class, and 340 were students in the laboratory of the study skills course; and the remaining 187 were in the reading improvement course.

The lecture evaluation revealed that 54 of the 72 members of the Spring, 1952, class were favorably disposed to the lectures; 50 were interested to the extent that they volunteered constructive suggestions to improve the lectures; and many could remember, and did write down the ways the lectures helped them. (There were 83 such responses.)

The laboratory evaluation, involving 344 students over a period of three successive semesters (Fall, 1951-52 through Fall, 1952-53), indicated that only 20 students (5.8%) felt they had received little or no help; 171 (49.7%) indicated that the assistance they received was either "considerable" or "great;" 143 (41.6%) felt they had received "some help;" only one felt that no help whatsoever had been received; 10 (2.9%) did not respond. Also 236 (68.6%) stated that they were encouraged to investigate and attempt to solve their study problems as against 88 (25.6%) who were not stimulated in this manner (5.8% did not volunteer comments in this area). 255 (71.2%) could remember, and did write down the laboratory periods which helped them most; and 131 (38.0%) could remember and did write down the laboratory periods which helped them least.

The reading course evaluation, involving 187 students over the Spring and Fall semesters of 1952-1953, indicated that only 9 (4.8%) felt they had received "little" help; while 122 (59.9%) stated that the help they had received was either "considerable" or "great;" and not one felt that he had received no help at all. 158 (84.5%) felt stimulated to investigate and attempt to solve their study problems, while 29 (15.5%) were not stimulated in this manner or did not volunteer this type of response. Significantly, 138 (73.8%) noted reading improvement in other university courses; a total of 49 (26.2%) either had no such experiences, or did not respond in this area.

In summary, it seems evident that the students who were assigned to a probationary status, and were required to participate in the University of Maryland study skills program, were benefited academically. Also, the great majority recognized and stated the benefits they felt they received despite the opportunity to be abusive which was open to every student in an anonymous evaluation of this sort. In view of the findings reviewed here it appears that the compulsory training given students does help many achieve their immediate goals despite the stigmatization suffered in varying degrees by students when first placed on probation.

Page 1
Form 3526
Rev. 8–50

Read the instructions on page 2 of this form.

STATEMENT REQUIRED BY THE ACT OF AUGUST 24, 1912, AS AMENDED BY THE ACTS OF MARCH 3, 1933, AND JULY 2, 1946 (Title 39, United States Code, Section 233) SHOWING THE OWNERSHIP, MANAGEMENT, AND CIRCULATION OF

JOURNAL OF EXPERIMENTAL EDUCATION published QUARTERLY
(Insert exact title of publication) (State exact frequency of issue)

at MADISON, WISCONSIN for October 1, 1956
(Name of post office and State where publication has second-class entry)

1. The names and addresses of the publisher, editor, managing editor, and business managers are:

	Name	*Address*
Publisher	DEMBAR PUBLICATIONS, INC.	303 E. WILSON STREET
Editor	A. S. Barr	Madison, Wisconsin
Managing editor	Jean W. Williams	Madison, Wisconsin
Business manager	Walter A. Frautschi	Madison, Wisconsin

2. The owner is: (If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual member, must be given.)

Name	*Address*
DEMBAR PUBLICATIONS, INC.	MADISON, WISCONSIN
A. S. BARR	MADISON, WISCONSIN
F. S. BRANDENBURG	MADISON, WISCONSIN
W. A. FRAUTSCHI	MADISON, WISCONSIN

3. The known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages, or other securities are: **(If there are none, so state.)**

Name	*Address*

4. Paragraphs 2 and 3 include, in cases where the stockholder or security holder appears upon the books of the company as trustee or in any other fiduciary relation, the name of the person or corporation for whom such trustee is acting; also the statements in the two paragraphs show the affiant's full knowledge and belief as to the circumstances and conditions under which stockholders and security holders who do not appear upon the books of the company as trustees, hold stock and securities in a capacity other than that of a bona fide owner.

5. The average number of copies of each issue of this publication sold or distributed, through the mails or otherwise, to paid subscribers during the 12 months preceding the date shown above was: (This information is required from daily, weekly, semiweekly, and triweekly newspapers only.) ____________

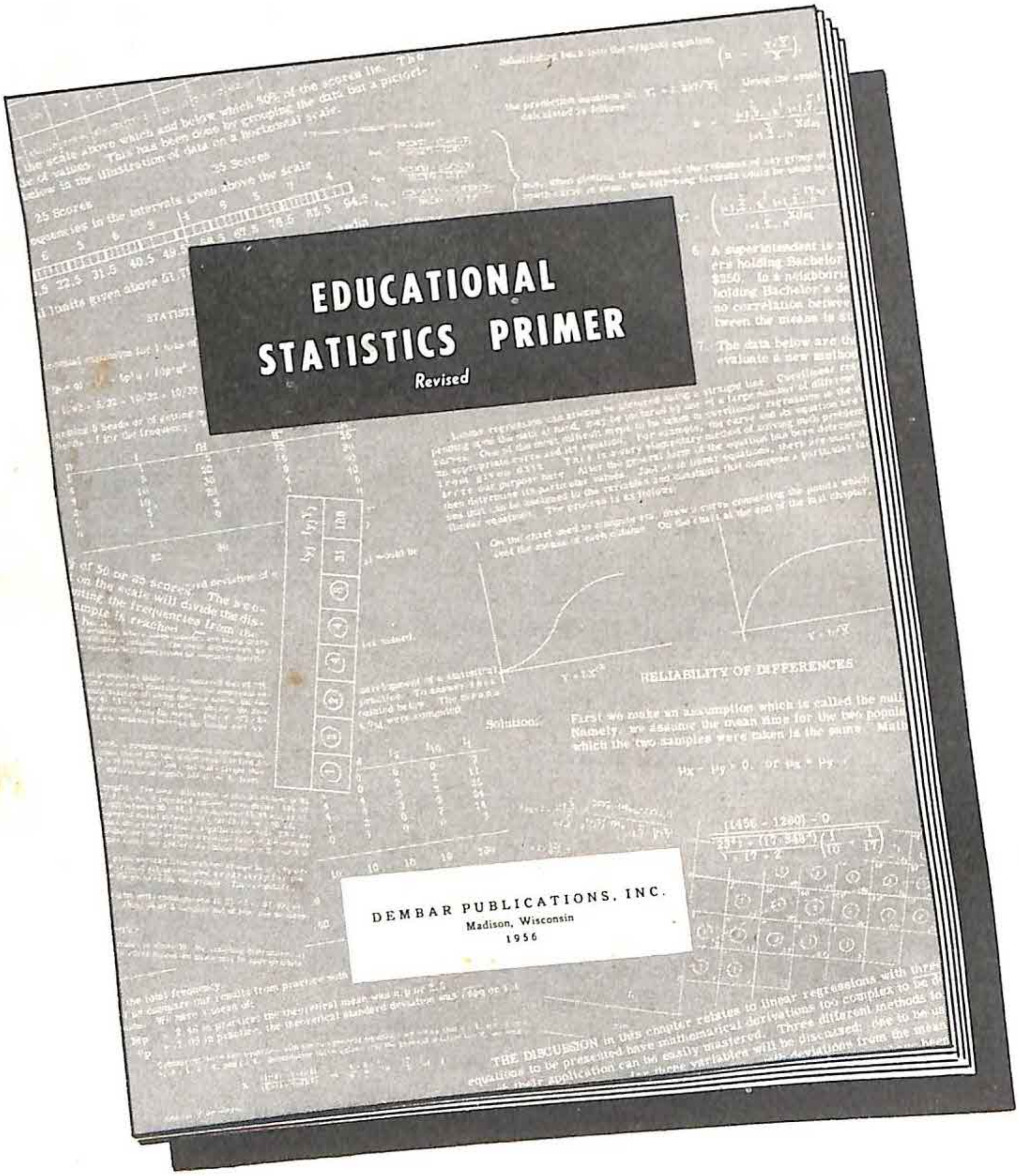
A. S. Barr, president
(Signature of editor, publisher, business manager, or owner)

Sworn to and subscribed before me this ____________ day of ____________, 19____

[SEAL] ____________

(My commission expires ____________, 19____)

16—18720-6



EDUCATIONAL STATISTICS PRIMER

Revised

DEMBAR PUBLICATIONS, INC.
Madison, Wisconsin
1956

Journal of Experimental Education

Volume XXV December, 1956 Number 2

A STUDY OF THE RELATIONSHIPS OF CERTAIN ENVIRONMENTAL FACTORS TO TEACHING SUCCESS

WILLIAM B. KNOX*
Platteville State Teachers College
Platteville, Wisconsin

SECTION I

INTRODUCTION

The Problem

THIS STUDY seeks to ascertain the relationships of certain environmental factors to teaching success. The environmental factors to be studied include the many conditions and facilities, present or absent, in a given situation that may influence the teacher's effectiveness. A potentially upsetting condition may be present or absent but a particular individual may, because of his temperament or philosophy of life, not be upset by it. Even though individuals vary greatly in adaptability to situations, it would seem desirable to ascertain the aspects of the environment to which teachers react unfavorably and express beliefs and attitudes which may influence their effectiveness.

Assuming that environmental factors expressed in an attitudinal manner may influence the effectiveness of a teacher, this study seeks to answer the following questions:

1. With what situational factors do teachers appear to be most satisfied or dissatisfied?
2. What situational factors seem to be associated with teaching success?
3. To what degree can teacher success be predicted on the basis of situational factors?
4. What situational factors do teachers think should be changed?

In order that we can effectively predict and control the behavior of people over a longer period of time it would seem advisable to know more about the situation in which people function. A better knowledge of the environment, and the individual's reactions to it, should lead to a higher degree of prediction and control.

It is common observation that some teaching situations are more complex than others. The recognition of the environment as an important aspect of living was pointed out some years ago by the sociologist W. I. Thomas (27). He indicated the areas in which the investigation of behavior problems may be profitably pursued, viz., "the attitudes, the values, the forms of adaptation, and the total situation."

In any study of human relations there is need of a psychology that provides adequate recognition of environmental factors. Such a psychology should give an adequate social consideration to relations and forces in a dynamic interactive situation. Lewin (16) concludes: "Every psychological event depends upon the state of the person and at the same time on the environment.... The experimental work of recent years shows more and more this twofold relationship in all fields of psychology. Every scientific psychology must take into account whole situations, i.e., the state both person and environment." Murphy (21) states, "It must be recognized that the slighting of situational differences has been a constant source of clinical and experimental errors; the clinical interview and the laboratory have both consistently failed to predict adequately the nature of social behavior that occurs in situations other than those in which the observations are originally made. The laboratory psychologist and the clinician have both conceived the individual as a system of events and tendencies carried around within the skin of the individual subject. They have both assumed that his delinquencies or his triumphs, his interests or his nervous maladjustments, result fully and simply from structured disposition within him." Murphy goes on to say that "a study of situations that act upon persons should be at least as full and as systematic as is a study of internal structures which respond to these situations."

Barr (3) indicates in a summary of recommendations for the study of the prediction of teaching efficiency that,

Bureau Ednl. Psy. Research
DAVID [illegible] COLLEGE
Dated
Acc. No.

---

*The author wishes to express his appreciation to Dr. A. S. Barr for his constant inspiration, guidance and encouragement during the course of his graduate study and the completion of this investigation.

> More attention should be given to the situational factors in success. It is easier to succeed in some situations than others. More attention needs to be given to these situational aspects of efficiency.

McClusky and Strayer (19), using a teacher situation test, came to the conclusion that "practically every phase of the teacher's life, ranging from his relation to the community to his attitude toward equipment in the classroom, is involved in his adjustment." Barr and Low (17) found a positive correlation between the complexity of teaching situations and teaching success. Again, in a summary of recommendations, Barr (3) concludes, "It is probably erroneous to assume that specific acts are good or bad divorced from purposes, principles, persons, and situations...."

It would seem that in light of these psychological observations and the attitudes expressed as to the importance of environmental factors in predicting behavior, that this problem is one of paramount importance.

Research Directly Related to the Problem

The paucity of research directly related to this area of investigation is probably due to the elusive character and the complexity of the elements that are the components of the environment.

It is important to know the personal characteristics of an individual. Such knowledge enables one to understand, at least in part, the actions of individuals in any situation. It is becoming increasingly recognized, however, that this alone is not sufficient.

Lins (17) studying 58 women teachers in Wisconsin schools, using a composite of five supervisory ratings, reports a composite evaluation of the situation with a correlation of -.34. Garrison (10) concluded after investigating job satisfaction of 46 women college seniors who had just finished 8 weeks of practice teaching and 42 elementary women teachers in service that, "Although the teacher-pupil relationship is of utmost importance in relation to satisfaction, many other situations in the teacher's life affect the teacher's feeling of satisfaction." Simon (21) after studying the records of action taken in the dismissal of 769 teachers and administrators concludes, "They were unable to make satisfactory adjustments. The teacher's inability to adjust himself to the community or to the school environment will cause him to be unpopular with the pupils and parents and may cause boys and girls to dislike the subjects he teaches." Symonds (26) had 65 teachers write a paper telling how they solved their personal problems. The three problems most frequently mentioned in their themes, in order of frequency, were: feelings of inadequacy and inferiority, professional difficulties, and family relations. Since living does not occur in a vacuum it might be well to know in such cases what the situational factors were that lead to the buildup of these problems. Martindale (18) gathered data from 72 teachers by means of a Teacher Situation Inventory that contained 316 situational items. He attempted to ascertain the factors that presumably lay outside the personality of the teacher. He maintained,

> As individual teachers differ in many ways so do all schools and communities differ. Some schools are progressive while others are conservative; some communities are friendly, others aloof; some are tolerant, others exacting. Population makeup, wealth, cultural status, and many other factors create differences in schools and communities. Every school and community is unique just as every individual person is unique.
>
> It is logical to suggest, then, that some of the factors influencing teaching success might possibly lie outside the teacher's control and within her... Possession of certain factors commonly associated with teaching success is not sufficient to predict success; factors which predict success might be conditioned by the teacher's compatability with her teaching environment.
>
> Numerous instances of an apparently well-trained and well-qualified teacher failing in one community while succeeding in another can be cited by superintendents and placement officials.

Martindale reported a correlation of .19 between teaching success and satisfaction scores.

The work that has been done in the area of the situational factors as related to teaching effectiveness is inconclusive. Many studies have been reported that place the emphasis on personal factors operating in some poorly understood environmental media. The concern of this investigation is to identify certain environmental factors which may influence the prediction of teaching success.

SECTION II

DESIGN OF THE STUDY

A DESCRIPTION of the subjects, the instruments used to collect the data, and proposals for analyzing the data will be presented in this section.

Description of the Subjects

This study is a descriptive investigation of the situational factors that may influence the effectiveness of teachers. The subjects of the investigation were sixty seniors who had graduated in 1950 from a four-year course in the School of Education at the University of Wisconsin. The group is divided into two sub-groups: one group of 45 was composed of 45 individuals who responded to a questionnaire; and a second group of 15 non-responding individuals who were interviewed. A summary of the number of majors and minors by fields of study for the group of 45 is given in Table I.

From an examination of the data in Table I it can be seen that there was a concentration in the fields of English and General Science and an even spread in the other subject fields. In the group of subjects used in this investigation there were 22 women and 23 men. The men had an average age of 25.7 years with a range of 22 to 30 inclusive. Seventeen of the men had military service. Only two of the women had military service. The women's average age was 23.2 with a range of 21 to 30 years. The average age for both men and women was 24.5 years.

The personal information secured relative to the subjects of this study showed the following: The men had 21 younger or older brothers with a range in number of 1 to 4. They also had 21 older or younger sisters with a range in number of 1 to 3. The women came from families with fewer younger and older brothers, but there were more older sisters.

It appears in the matter of residence that the families of these persons were relatively stable and moved only within the confines of their community. The number of schools attended, the occupation of the parents, and population centers, in which they lived at one time, indicate this fact. Table II indicates the number of schools attended.

The subjects of this investigation lived in various sized communities as indicated in Table IV.

In the number of jobs held for pay the women worked most frequently in the clerical and sales areas. The men performed manual labor, and mechanical and factory work with nineteen having experience in the professional and technical fields.

As to the leadership experiences, the men held as many class officerships as women, the latter had more experiences as Sunday school teachers, tutoring others and serving on councils and committees. The men had most of their experience as camp counselors, Scouts, 4-H Club leaders with a scattering of Sunday School teaching and tutoring. While attending college there were nine men and two women who were self-supporting. Eleven of the men and five women were married at the time of this investigation.

Comparative data for the members of this study and a group studied by Lins (17) and Montross (20) are given in Table V. The grade point average for the required Education courses 73, 74, and 75, which includes practice teaching is given as well as the total four-year grade point average for various groups.

A follow-up interview was made of fifteen of the subjects of this investigation who did not respond to the Situation Inventory. In this group, selected at random, were four men and eleven women with major fields of study as follows: 4 in English, 3 in Physical Education, 2 in Art Education, 1 in Physics, 2 in Natural Science, 1 in Home Economics, 1 in Speech Correction, and 1 in History. Their minor fields were French, Spanish, Political Science, General Science, Mathematics, English, Botany, Zoology, Chemistry, Sociology, and Speech.

The mean grade point for Education 73, 74, and 75, and the four-year grade point average for the sample did not differ markedly from the larger group of subjects studied (Table VI).

It appears, from the data given in Tables V and VI that the subjects of this investigation, are representative of University of Wisconsin graduates.

Description of the Data-Gathering Devices

A. Teacher Situation Inventory

The Inventory (Appendix A-1*) used in this investigation was, in substance, the same as the one devised and used by Frank Martindale in his study of Situational Factors in Teacher Placement and Success (18). The directions for filling out the inventory were extensively revised, however, by the present investigator to facilitate the logical step by step procedure in completing the inventory. The format for the write-in space was also changed.

The revised inventory consisted of two pages of directions and nine sections of items listed under the following categories: the community, the people of the community, the school plant, the school administration, the school organization, the instructional materials, the faculty, the students, and the particular position. These nine categories made a total of 317 items. The aim of the inventory was to determine the degree of satisfaction and dissatisfaction of each respondent with respect to the several items and the reason for his responses. The last page of the inventory consisted of a request for general information.

B. How the Data Were Collected

The data for this investigation were secured

---

*All references to Appendices may be found in original thesis, Library, University of Wisconsin.

TABLE I

MAJORS AND MINORS

Field of Study	Men		Women		Total	
	Major	Minor	Major	Minor	Major	Minor
American Institutions	3				3	0
Art	3		1		4	0
Business Education	2		2		4	0
Chemistry	1				1	0
Economics		4		2	0	6
English	2	6	6	6	8	12
General Science		6	3	5	3	11
History	4	4			4	4
Home Economics			4		4	0
Languages			2	6	2	6
Mathematics		3	1	1	1	4
Music	1				1	0
Physical Education	4				4	0
Speech Correction	1		3		4	0
Others	2			1	2	1
Total	23		22		45	

TABLE II

NUMBER OF SCHOOLS ATTENDED

Type of School	Men				Women			
	Number of Schools Attended				Number of Schools Attended			
	1	2	3	4	1	2	3	4
Elementary	12	5	5	0	19	4	0	0
Secondary	19	3	0	0	20	3	0	0
College	16	4	1	1	18	5	0	0
Trade	4	0	0	0	1	0	0	0

TABLE III

OCCUPATIONAL STATUS OF WOMEN

Men	Women
Agent, Freight	Accountant
Agent, Purchasing	Baker
Assembler	Banker
Attendant, Filling Station	Bus Line and Transfer
Baker	Buyer
Buyer, Cheese	Clerk, Store
Clerk, Railroad	Dealer, Electrical-Plumbing
Farmer (3)*	Farmer (4)
Foreman, Brass Factory	Forewoman, Factory
Foreman, Paper Mill	Gardner
Grocer (2)	Janitress (3)
Investment	Manager, Office
Janitor (1)	Physician
Lawyer	Professor, University
Merchant	Saleslady, Auto
Minister	Scrap Metal Dealer
Oil Dealer	Superintendent of Schools
Postmaster	Trucker
Store Owner	
Total. 22	Total. 23

*Indicates the number in that category.

TABLE IV

NUMBER AND AVERAGE YEARS LIVED AT SOME TIME IN THEIR LIVES

Size of Community	Men No.	Men Ave.	Women No.	Women Ave.
Farm	4	10	4	17
Town under 500	4	10	2	21
500 to 2500	7	13	5	12
2500 to 10,000	8	15	6	10
10,000 to 100,000	9	11	12	15
Over 100,000	6	17	4	13

TABLE V

A COMPARISON OF MEAN GRADE POINT AVERAGES

Measure	Mean 1	Mean 2	Mean 3
Education 73	2.18	2.31	2.34
Education 74	2.13	2.13	2.02
Education 75	1.90	2.37	2.17
Total Four Year Grade Point Average	1.84	1.93	1.88
Total Four Year G. P. A. of 586 Education Seniors who graduated in 1950	2.02		
Total Four Year G. P. A. of 4577 all University graduating seniors	1.88		

Key

1. Subjects of this investigation
2. Study by Montross
3. Study by Lins

TABLE VI

MEAN GRADE POINT FOR THE SAMPLE IN COMPARISON WITH THE SUBJECTS OF THIS INVESTIGATION

Measure	Mean 1	Mean 2
Education 73	2.06	2.18
Education 74	2.33	2.13
Education 75	2.06	1.90
Total Four Year Grade Point Average	1.98	1.84
Total Four Year G.P.A. of 586 Education Seniors who graduated in 1950	2.02	
Total Four Year G.P.A. of 4577 all University Graduating Seniors	1.88	

Key

1. Fifteen subjects of the sample
2. All subjects of the investigation

TABLE VII

ORIGINAL SCALE AND WEIGHTING SCALE

1	+2	extremely satisfactory, highly agreeable
2	+1	mild satisfaction or agreeable
3	0	neutral, not affecting liking or disliking
4	-1	mildly unsatisfactory, disagreeable
5	-2	extremely unsatisfactory, highly disagreeable

from a follow-up study of sixty-five 1950 graduates of the School of Education at the University of Wisconsin in their first year of teaching. The Teacher Situation Inventory (Appendix A-1) was sent by mail to each member of this group during the last month of the first year of his teaching experience. A letter inviting continued cooperation was sent with the inventory (Appendix A-2). Two follow-up letters were sent subsequently to those who were tardy in returning the inventory (Appendix A-3 and A-4). Forty-nine returned a filled-out inventory. Of those returned, four were not usable because of the omissions and inadequacy of write-in replies as to why they expressed a degree of satisfaction or dissatisfaction.

C. The Criteria of Teaching Success

The criteria used in this investigation were as follows:

1. Acceptability Rating

   This is a rating used in an investigation by Lamke[1] and Ringness[2]. It was obtained by interviewing the principal or superintendent. After rapport had been established the following questions were asked to obtain his overall rating of the teacher:

   a. How do you like this teacher?
   b. Does she get along well with the students?
   c. Do the parents feel the same way as the pupils do about the teacher?
   d. Is the teacher happy here? Does she have friends on the faculty, in the community, and the like?
   e. Does the teacher have any particular weaknesses or strengths?
   f. In terms of beginning teachers with whom you have worked in the past, and in the light of our discussion as a whole, would you consider this teacher as among the best, above average, average, below average, or failing as a teacher?

   The above subjective rating thus obtained from a supervisory or administrative official was defined by Ringness[3] as, "... that quality of the teacher which leads other people to feel that she should be retained in position or possibly be promoted to more responsible assignments."

2. A short form of the Wisconsin Adaptation of the M-Blank (Appendix A-5) was used to obtain the principals' or superintendents' estimate of the teachers' performance. The Adaptation, which was the work of Barr and others is a modification of the M-Blank of the Evaluative Criteria. This Blank in this study included:

   a. Elements of Strength in the Teacher's Work
   b. Elements of Weakness in the Teacher's Work
   c. General Evaluation: Outstanding, Above Average, Average, Below Average, Poor
   d. General Comment

3. A peer rating scale (Appendix A-6) was sent to a co-worker who was acquainted with the particular subjects' work. It contained the following:

   a. How do you rate this teacher on cooperation with you in school enterprises?
   b. Do you find this teacher pleasant and agreeable?
   c. How do you rate this teacher on social intelligence?
   d. What is your opinion concerning this teacher's physical vitality and health?
   e. What is your opinion concerning this teacher with respect to industry and dependability?
   f. To what extent does this teacher participate in community affairs?
   g. How well does this teacher know his subject matter fields?
   h. How do you classify his relationships with the student body?
   i. You are asked to make an overall evaluation of this teacher. If you were to rank all teachers of similar positions, in which fifth would you place this teacher?

4. The judgment of a supervisory official of the State Department of Public Instruction and designated as criterion (X_1).

These criteria were used singly and as a composite of the four measures. In addition, where the subjects of this investigation overlapped with those of Montross (7), his "Number One" composite criterion was used. This "Number One" composite included the Principal's Acceptability rating, the Principal's M-Blank rating, Rater 1 and Rater 2.

---

[1] Tom Arthur Lamke. Personality and Teaching Success, unpublished manuscript, University of Wisconsin, Madison, Wisconsin, 1951.

[2] Thomas A. Ringness. Relationships Between Certain Attitudes Towards Teaching and Teaching Success, Unpublished Ph.D. Dissertation, University of Wisconsin, 1951.

[3] Ibid., p. 9.

The effectiveness of each teacher was measured through the use of these instruments as guides to the judgment of the rater.

D. Treatment of the Data

The data are tabulated and the mean frequency of the response to each item tabulated. The mean scores designating satisfaction and dissatisfaction were correlated with efficiency scores on the above named criteria. Selected items of the Inventory, that show low dissatisfaction and high satisfaction scores with a summary of responses given by the subjects of this investigation for such items, are reported.

SECTION III

ANALYSIS OF DATA

THE PURPOSE of this study is to ascertain the relation of certain environmental factors to teaching success as measured by the efficiency scores in this investigation.

The primary sources of data were the teaching efficiency ratings assigned by experts and the responses of the teachers to the Teacher Situation Inventory. Respondees checked each item of the Inventory using a +2, +1, 0, -1, and -2 scale to indicate degrees of satisfaction and dissatisfaction. This scale was retained in computing totals, also the plus, minus and neutral responses were assigned values from 1 to 5 as shown in Table VII. Using the weighted scores facilitated the handling of grouped data; retaining the algebraic scores facilitated rapid identification of items to which respondees indicated dissatisfaction and identification of individual teachers who indicated dissatisfaction.

The response to each item of the inventory by each respondent was recorded as shown in Appendix B-1. The responses of each individual were entered in a working chart of which Table VIII is an example. In this table the responses to items in the nine categories of the inventory are listed. Both methods for scoring items were employed to arrive at a total weighted score for each category. A mean weighted score for each category, a total algebraic score for each category, a total weighted score for all categories, a mean weighted score for all categories, and a total algebraic score for all categories was calculated.

The total mean weighted score for all categories, the mean weighted score for each category, and the algebraic score for each category of each respondee were entered into a summary table (Table IX).

The summarized inventory information was then transferred to a final master chart, Table X. In this table is recorded each respondee number, sex, four teaching efficiency ratings, a peer rating, a composite of three efficiency ratings and a peer rating. Finally, the satisfaction-dissatisfaction scores were converted into percents. The percent satisfaction-dissatisfaction score was obtained by subtracting 3.00 from the mean score and dividing by 2 when the mean score exceeded 3.00, and by subtracting the mean score from 3.00 and dividing by 2 when the mean score was less than 3.00.

It should be noted in the method of scoring explained above that the teachers may respond negatively, positively, or neutrally to any given item. If the teacher responded favorably he was given a plus score; if he responded negatively he was given a negative score. If he responded neutrally he was given a zero score. It should be further noted that each teacher may respond either positively or negatively in two degrees of intensity, namely, minus 1 or minus 2, negatively, and plus 1 or plus 2, positively. When one adds such a series of scores the final sum will have a plus or minus sign which represents a mixture of frequency and intensity, i.e., a teacher may get a negative score by responding negatively to a large number of items mildly (-1) or to a lesser number of items strongly (-2). The same explanation applies to the positive responses.

When we speak of a percent of satisfaction or dissatisfaction we are speaking of an abstract value derived by dividing by 2 the theoretical mean for all items (3.00) minus the particular teachers' mean upon the particular items concerned.

One hundred percent satisfaction means maximum satisfaction with all items. All departures from neutrality are expressed by a fraction, the numerator of which is the deviation of each individual's score from 3.00 over a denominator of 2. This fraction is used to represent the maximum possible deviation. A plus score of 2 represents complete satisfaction to any given item and will be represented by the fraction 2/2 or one hundred percent. The satisfaction of an individual with a plus score of 1, will be represented by the fraction 1/2 or fifty percent.

The reader should further note that each individual's total response is here represented by an average which is secured by dividing the sum of the scores by the number of responses. If the average is numerically less than 3 it is interpreted as a satisfaction score. If it is more than 3, it is treated as a negative score, and the percentages are interpreted accordingly. Possibly attention again should be called to the fact that a small plus or minus score may arise from strong feeling about a relatively few items. The balance may be changed from plus to minus or minus to plus by a few of these items when the

Bureau Ednl. Psy. Research
DAVID [illegible] COLLEGE
Dated
Acs No

TABLE IX

MEAN WEIGHTED AND ALGEBRAIC SCORES OBTAINED FROM RESPECTIVE CATEGORIES OF THE TEACHER SITUATION INVENTORY

Subject Number	I The Community		II People of Community		III School Plant		IV Administration		V School Organization		VI Instructional Materials		VII Faculty		VIII Students		IX Particular Position		Mean Total Percent
	(a)**	(b)***	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	
1	3.04	- 3*	3.29	- 6	3.59	-19	1.94	+19	2.87	+ 5	2.64	+ 9	2.31	+31	2.21	+19	2.53	+17	2.77
2	2.09	+63	3.33	- 7	1.88	+36	1.78	+22	1.84	43	2.32	18	1.73	57	2.46	13	2.26	47	2.11
3	1.70	73	3.14	- 3	3.92	-25	1.33	15	2.63	10	2.65	7	2.57	9	4.33	- 4	2.52	11	2.55
4	1.44	76	1.48	14	1.34	53	1.50	24	1.90	43	1.60	35	1.87	36	1.00	33	2.61	14	1.67
5	2.50	35	3.05	- 1	2.78	7	2.11	16	2.62	12	2.62	10	2.27	33	3.00	0	2.06	29	2.53
6	1.84	77	2.48	11	2.63	12	1.00	36	1.15	72	1.00	50	1.67	60	2.92	2	1.58	51	1.79
7	3.16	-11	3.24	- 5	3.53	-17	3.33	- 6	3.26	-10	3.23	- 6	3.22	-10	3.38	- 9	3.28	-10	3.27
8	3.00	0	3.19	- 4	3.25	- 8	2.50	9	2.77	9	2.40	15	2.69	14	2.83	4	1.89	40	2.74
9	1.77	86	1.71	27	2.23	23	2.00	18	2.90	4	2.48	13	2.13	39	1.71	31	1.42	49	2.04
10	2.61	18	2.86	3	2.66	11	1.61	25	2.05	31	2.72	7	1.80	54	3.04	- 1	2.50	12	2.41
11	2.71	20	3.35	- 7	2.75	4	2.47	9	3.03	- 1	3.88	-22	3.77	-17	3.67	-14	3.36	-10	3.15
12	2.16	58	3.05	- 1	2.78	7	1.67	24	2.46	21	1.96	26	1.98	46	2.92	2	1.67	48	2.25
13	1.72	88	2.90	2	1.50	48	1.44	28	1.33	65	1.16	46	1.43	69	1.18	40	1.18	62	1.53
14	1.97	72	1.76	26	2.16	27	1.00	36	1.44	61	1.52	37	1.49	68	1.57	33	1.71	44	1.68
15	3.24	-17	3.19	- 4	3.90	-29	3.89	-16	3.54	-17	4.32	-33	2.58	19	4.08	-26	3.08	- 3	3.42
16	3.01	- 1	3.00	0	2.32	21	2.17	15	2.59	16	3.56	-14	3.04	- 2	3.25	- 6	2.56	16	2.85
17	2.78	15	3.10	- 2	3.94	-30	3.56	- 9	3.08	- 3	3.44	-11	2.18	36	3.13	- 3	3.28	-10	3.06
18	2.36	44	2.90	2	3.75	-24	2.67	6	3.67	-26	2.96	1	2.71	13	3.33	- 8	2.69	11	2.94
19	3.09	- 6	3.38	- 8	2.56	14	3.38	- 6	3.00	0	2.65	9	2.96	2	3.29	- 7	2.58	15	2.96
20	2.32	44	2.24	16	1.53	47	1.83	21	1.90	43	2.00	24	2.20	33	2.08	22	1.94	36	1.96
21	2.03	67	1.62	29	1.06	62	1.00	36	1.21	70	1.09	44	1.36	74	1.83	27	1.36	59	1.47
22	2.87	9	2.52	10	3.25	- 8	2.50	9	3.23	- 9	3.23	- 6	3.00	0	2.96	1	3.08	- 3	2.99
23	2.32	42	2.43	12	1.31	54	1.35	28	1.82	46	1.78	28	1.80	53	2.28	13	1.97	34	1.94
24	3.21	-15	3.1	- 2	2.34	21	2.78	4	2.72	11	2.96	1	3.04	- 2	2.92	2	2.94	2	2.93
25	2.67	23	2.57	9	3.90	- 9	2.33	10	2.74	10	2.62	10	1.78	45	1.64	30	1.92	39	2.39
26	3.51	-36	3.29	- 6	3.41	-13	2.75	3	3.28	-11	2.48	13	3.02	- 1	3.13	- 3	3.00	0	3.18
27	1.97	72	1.43	33	2.16	27	1.50	27	1.87	44	1.65	35	1.67	60	2.25	18	1.67	48	1.83
28	2.17	58	2.24	16	2.06	30	1.67	24	1.92	42	1.31	44	1.52	65	2.83	4	1.81	43	1.95
29	3.03	- 2	2.81	4	2.19	26	2.71	5	2.43	21	2.96	1	2.00	44	2.25	18	2.78	8	2.59
30	2.41	37	2.05	20	2.75	8	2.56	8	2.54	18	2.79	5	2.36	29	3.21	- 5	3.02	- 1	2.61
31	2.74	18	2.71	6	1.91	35	1.78	22	2.44	22	2.88	3	2.15	34	3.17	- 4	1.97	29	2.43
32	1.87	76	2.16	16	1.22	57	1.44	28	1.28	67	1.17	42	1.79	41	1.24	37	1.57	37	1.56
33	2.28	49	2.62	8	1.09	61	1.06	35	1.69	51	1.92	28	2.00	45	1.38	39	1.81	43	1.84
34	2.19	56	1.52	31	2.72	9	2.00	18	1.62	54	1.84	29	1.31	76	1.75	30	1.86	41	1.89
35	2.83	12	2.95	1	1.94	34	1.94	19	2.44	22	2.96	1	2.49	23	3.25	-14	269	11	2.62
36	1.59	97	1.90	23	1.97	43	1.11	34	3.08	- 3	2.28	18	2.49	23	2.75	6	2.78	8	2.23
37	2.43	33	3.10	- 2	2.91	3	2.17	15	2.77	9	2.68	8	2.55	20	3.13	- 3	2.86	5	2.70
38	2.03	66	2.00	21	2.00	26	3.00	0	2.67	12	3.50	-13	3.27	- 9	2.75	4	2.83	2	2.55
39	2.10	63	2.95	1	2.78	7	1.44	28	1.97	40	2.05	20	1.73	57	2.75	6	1.29	48	2.09
40	2.52	33	2.29	15	1.88	36	2.22	14	2.15	33	2.24	19	2.29	32	2.21	19	2.42	21	2.28
41	1.73	89	2.05	20	2.03	31	2.56	8	2.79	8	2.96	1	2.75	11	2.75	6	2.58	15	2.39
42	2.39	42	2.90	2	2.53	15	2.00	18	2.31	27	3.24	- 6	2.39	25	2.71	7	2.52	14	2.52
43	2.88	8	2.90	2	3.88	-25	1.44	28	1.97	30	2.48	13	1.62	62	2.54	11	2.44	20	2.50
44	1.56	101	2.00	21	3.09	- 3	1.83	21	2.56	17	2.50	13	2.11	40	2.29	17	2.47	19	2.21
45	2.09	64	2.38	12	2.53	15	1.44	28	1.46	60	2.20	20	1.93	48	2.71	7	1.47	55	2.00

* Minus signs indicate negative responses.
** Weighted Mean Score 1 - 5
***Weighted Algebraic Score

TABLE VIII

RAW SCORES OBTAINED FROM INVENTORY (WORK SHEET)

Weighted Value	Algebraic Value	I No. of items related to community			II No. of items related to people of community			III No. of items related to school plant			IV No. of items related to administration			V No. of items related to school organization			VI No. of items related to instructional materials			VII No. of items related to faculty			VIII No. of items related to students			IX No. of items related to particular position			
			(a)*	(b)**		(a)	(b)		(a)	(b)		(a)	(b)		(a)	(b)		(a)	(b)		(a)	(b)		(a)	(b)		(a)	(b)	
1	+2	10	10	20	2	2	4	1	1	2	4	4	8	8	8	16	4	4	8	15	15	30	4	4	8	6	6	12	
2	+1	14	28	14	0	0	0	8	16	8	12	24	12	10	20	10	9	18	9	16	32	16	15	30	15	15	30	15	
3	0	24	72	0	10	30	0	1	3	0	1	3	0	0	0	0	4	12	0	5	15	0	1	3	0	6	18	0	
4	-1	7	28	- 7	8	32	- 8	15	60-15		1	4	- 1	21	84	-21	8	32	- 8	3	12	- 3	4	16	- 4	8	32	- 8	
5	-2	15	75	-30	1	5	- 2	7	35	-14	0	0	- 0	0	0	0	0	0	0	6	30	12	0	0	0	1	5	- 2	
Total items responded to:		70			21			32			18			39			25			45			24			36			310
Total weighted score			213			69			115			35			112			66			104			53			91		858
Total original score				- 3			- 6			-19			19			5			9			31			19			17	+72
Mean			3.04			3.29			3.59			1.94			2.87			2.64			2.31			2.21			2.53		2.77

* (a) Weighted scores
**(b) Algebraic scores

TABLE X

MASTER CHART

Subject Number	Sex	Montross' Composite Rating	Principal Acceptability	M-Blank	Criterion (X1)	Peer Rating	Composite (X0)	Community		People of Community		School Plant		Administration		School Organization		Instructional Materials		Faculty		Students		Particular Position		Mean Total	Total Mean Percent	Subject Number
		X_4	X_3	X_2	X_1		X_0	Y_1	Y_1	Y_2	Y_2	Y_3	Y_3	Y_4	Y_4	Y_5	Y_5	Y_6	Y_6	Y_7	Y_7	Y_8	Y_8	Y_9	Y_9	Y_0		
1	F	4.66	5	4	-	-	4.50	3.04	- 2.0	3.29	-14.5	3.59	-29.5	1.94	53.0	2.78	11.0	2.64	18.0	2.31	34.5	2.21	39.5	2.53	23.5	2.77	11.5	1
2	M	----	3	2	5	5	3.75	2.09	35.5	3.33	-16.5	1.88	56.0	1.78	61.0	1.84	58.0	2.32	34.0	1.73	63.5	2.46	27.0	2.26	37.0	2.11	44.5	2
3	M	3.00	2	3	1	-	2.00	1.70	65.0	3.14	- 7.0	3.92	-46.0	1.33	83.5	2.63	18.5	2.65	17.5	2.57	21.5	4.33	-33.5	2.52	24.0	2.55	22.5	3
4	M	----	2	1	1	5	2.25	1.44	78.0	1.48	76.0	1.34	83.0	1.50	75.0	1.90	55.0	1.60	70.0	1.87	56.5	1.00	100.0	2.61	19.5	1.67	66.5	4
5	M	1.00	1	1	1	-	1.00	2.50	25.0	3.05	- 2.5	2.78	11.0	2.11	94.5	2.62	19.0	2.62	19.0	2.27	36.5	3.00	0.0	2.06	47.0	2.53	23.5	5
6	M	2.00	2	2	2	2	2.00	1.84	58.0	2.48	26.0	2.63	18.5	1.00	100.0	1.15	92.5	1.00	100.0	1.67	66.5	2.92	4.0	1.58	71.0	1.79	60.5	6
7	M	----	3	3	5	3	3.50	3.16	- 8.0	3.24	-12.0	3.53	-26.5	3.33	-16.5	3.26	-13.0	3.23	-11.5	3.22	-11.0	3.38	-19.0	3.28	-14.0	3.27	-13.5	7
8	F	----	2	1	2	6	2.75	3.00	0.0	3.19	- 9.5	3.25	-12.5	2.50	25.0	2.77	11.5	2.40	30.0	2.69	15.5	2.83	8.5	1.89	55.5	2.74	13.0	8
9	M	2.25	2	2	-	-	2.00	1.77	61.5	1.71	64.5	2.23	38.5	2.00	50.0	2.90	5.0	2.48	26.0	2.13	43.5	1.71	64.5	1.42	78.0	2.04	48.0	9
10	F	3.50	4	2	6	-	4.67	2.61	19.5	2.86	7.0	2.66	17.0	1.61	69.5	2.05	47.5	2.72	14.0	1.80	60.0	3.04	- 2.0	2.50	25.0	2.41	29.5	10
11	M	----	1	2	-	-	1.50	2.71	14.5	3.35	-17.5	2.75	12.5	2.47	26.5	3.03	- 1.5	3.88	-44.0	3.77	-38.5	3.67	-33.5	3.36	-18.0	3.15	- 7.5	11
12	F	1.75	2	2	-	-	2.00	2.16	42.0	3.05	- 2.5	2.78	11.0	1.67	66.5	2.46	27.0	1.96	52.0	1.98	51.0	2.92	4.0	1.67	68.5	2.25	37.5	12
13	M	----	3	2	1	2	2.00	1.72	64.0	2.90	5.0	1.50	75.0	1.44	78.0	1.33	83.5	1.16	92.0	1.43	78.5	1.18	91.0	1.18	91.0	1.53	73.5	13
14	M	----	2	3	4	3	3.00	1.97	51.5	1.76	62.0	2.16	42.0	1.00	100.0	1.44	78.5	1.52	74.0	1.49	75.5	1.57	71.5	1.71	64.5	1.68	64.0	14
15	M	3.25	3	3	4	2	3.00	3.24	-12.0	3.19	- 9.5	3.90	-45.0	3.89	-44.5	3.54	-27.0	4.32	-66.0	2.58	21.0	4.08	-54.0	3.08	- 4.0	3.42	-21.0	15
16	F	2.00	2	2	3	3	2.50	3.01	- 0.5	3.00	0.0	2.32	34.0	2.17	41.5	2.59	20.5	3.56	-28.0	3.04	- 2.0	3.25	-12.5	2.58	22.0	2.85	7.5	16
17	F	----	1	2	1	3	1.75	2.78	11.0	3.10	- 5.0	3.94	-47.0	3.56	-28.0	3.08	- 4.0	3.44	-22.0	2.18	41.0	3.13	- 6.5	3.28	-14.0	3.06	- 3.0	17
18	F	2.75	3	3	3	2	2.75	2.36	32.0	2.90	5.0	3.75	-37.5	2.67	16.5	3.67	-33.5	2.96	2.0	2.71	14.5	3.33	-16.5	2.69	15.5	2.94	3.0	18
19	M	----	2	2	3	4	2.75	3.09	- 4.5	3.38	-19.0	2.56	22.0	3.38	-19.0	3.00	0.0	2.65	17.5	2.96	2.0	3.29	-14.5	2.58	21.0	2.96	2.0	19
20	M	2.50	2	2	1	5	2.50	2.32	34.0	2.24	38.0	1.53	73.5	1.83	58.5	1.90	55.0	2.00	50.0	2.20	40.0	2.08	46.0	1.94	53.0	1.96	52.0	20
21	F	2.00	2	2	3	6	2.75	2.03	48.5	1.62	69.0	1.06	97.0	1.00	100.0	1.21	89.5	1.09	95.5	1.36	82.0	1.83	58.5	1.36	82.0	1.47	76.5	21
22	M	3.25	3	2	2	3	2.50	2.87	56.5	2.52	24.0	3.25	-12.5	2.50	25.0	3.23	-11.5	3.23	-11.5	3.00	0.0	2.96	2.0	3.08	- 4.0	2.99	.5	22
23	F	2.25	2	3	4	4	3.25	2.32	34.0	2.43	28.5	1.31	84.5	1.35	82.5	1.82	59.0	1.78	61.0	1.80	60.0	2.38	31.0	1.97	51.5	1.94	53.0	23
24	M	----	2	2	2	2	2.00	3.21	-10.5	3.10	- 5.0	2.34	33.0	2.78	11.0	2.72	14.0	2.96	2.0	3.04	- 2.0	2.92	4.0	2.94	3.0	2.93	3.5	24
25	M	1.75	2	2	4	4	3.00	2.67	16.5	2.57	21.5	3.90	-45.0	2.33	33.5	2.74	13.0	2.62	19.0	1.78	61.0	1.64	68.0	1.92	54.0	2.39	30.5	25
26	M	1.66	1	1	2	4	2.00	3.51	-25.5	3.29	-14.5	3.41	-20.5	2.75	12.5	3.28	-14.0	2.48	26.0	3.02	- 1.0	3.13	- 6.5	3.00	0.0	3.18	- 9.0	26
27	M	----	2	2	3	3	2.50	1.97	51,5	1.43	78.5	2.16	42.0	1.50	75.0	1.87	56.5	1.65	67.5	1.67	66.5	2.25	37.5	1.67	66.5	1.83	58.5	27
28	M	3.00	2	2	4	6	3.50	2.17	41.5	2.24	38.0	2.06	47.0	1.67	66.5	1.92	54.0	1.31	84.5	1.52	74.0	2.83	8.5	1.81	59.5	1.95	52.5	28
29	M	----	2	3	1	5	2.75	3.03	- 1.5	2.81	10.5	2.19	40.5	2.71	14.5	2.43	28.5	2.96	2.0	2.00	50.0	2.25	37.5	2.78	11.0	2.59	20.5	29
30	F	----	2	3	6	5	4.00	2.41	29.5	2.05	47.5	2.75	12.5	2.56	22.0	2.54	23.0	2.79	10.5	2.36	32.0	3.21	-10.5	3.03	- 1.5	2.61	19.5	30
31	M	2.00	1	1	1	2	1.25	2.74	13.0	2.71	14.5	1.91	545	1.78	61.0	2.44	28.0	2.88	6.0	2.15	42.5	3.17	- 8.5	1.97	51.3	2.43	28.5	31
32	F	3.00	3	2	4	3	3.00	1.87	56.5	2.16	42.0	1.22	89.0	1.44	78.0	1.28	86.0	1.17	91.5	1.79	60.5	1.24	88.0	1.57	71.5	1.56	72.0	32
33	M	----	2	3	-	-	2.50	2.28	36.0	2.62	19.0	1.09	95.5	1.06	97.0	1.69	65.5	1.92	54.0	2.00	50.0	1.38	81.0	1.81	50.0	1.84	58.0	33
34	F	----	2	2	2	5	2.75	2.19	40.5	1.52	74.0	2.72	14.0	2.00	50.0	1.62	69.0	1.84	58.0	1.31	84.5	1.75	62.5	1.86	57.0	1.89	55.5	34
35	F	----	3	2	1	2	2.00	2.83	8.5	2.95	2.5	1.94	53.0	1.94	53.0	2.44	28.0	2.96	2.0	2.49	25.5	3.25	-12.5	2.69	16.5	2.62	19.0	35
36	F	----	3	3	2	5	3.25	1.59	70.5	1.90	55.0	1.97	51.5	1.11	94.5	3.08	- 4.0	2.28	36.0	2.49	25.5	2.75	12.5	2.78	11.0	2.23	38.5	36
37	F	----	2	3	5	6	4.00	2.43	28.5	3.10	- 5.0	2.91	4.5	2.17	41.5	2.77	11.5	2.68	16.0	2.55	22.5	3.13	- 6.5	2.86	7.0	2.70	15.0	37
38	F	3.75	4	3	-	-	3.50	2.03	48.5	2.00	50.0	2.00	50.0	3.00	0.0	2.67	16.5	3.50	-25.0	3.27	-13.5	2.75	12.5	2.83	8.5	2.55	22.5	38
39	F	----	3	2	-	-	2.50	3.10	45.0	2.95	2.5	2.78	11.0	1.44	78.0	1.97	51.5	2.05	47.5	1.73	63.5	2.75	12.5	1.29	85.5	2.09	45.5	39
40	F	----	2	2	-	-	2.00	2.52	24.0	2.29	35.5	1.88	56.0	2.22	39.0	2.15	42.5	2.24	38.0	2.29	35.5	2.21	39.5	2.42	29.0	2.28	36.0	40
41	F	1.33	1	3	1	-	1.67	1.73	63.5	2.05	47.5	2.03	48.5	2.56	22.0	2.79	10.5	2.96	2.0	2.75	12.5	2.75	12.5	2.58	21.0	2.39	30.5	41
42	F	2.50	2	2	5	4	3.25	2.39	30.5	2.90	5.0	2.53	23.5	2.00	50.0	2.31	34.5	3.24	-12.0	2.39	30.5	2.71	14.0	2.52	24.0	2.52	24.0	42
43	F	----	2	2	-	-	2.00	2.88	6.5	2.90	5.0	3.88	-44.0	1.44	78.0	1.97	51.5	2.48	26.0	1.62	69.0	2.54	23.0	2.44	28.0	2.50	25.0	43
44	M	2.00	2	2	1	3	2.00	1.56	72.0	2.00	50.0	3.09	- 4.5	1.83	58.5	2.56	22.0	2.50	25.0	2.11	44.5	2.29	35.5	2.47	26.5	2.21	39.5	44
45	F	1.00	1	1	1	3	1.50	2.09	45.5	2.38	31.0	2.53	26.5	1.44	78.0	1.46	77.0	2.20	40.0	1.93	53.5	2.71	14.5	1.47	76.5	2.00	50.0	45
Mean								2.39		2.40		2.60		2.53		2.04		2.38		2.46		2.24		2.63		2.31		
Percent									30.5		30.0		20.0		23.5		48.0		31.0		27.0		38.0		18.5		34.5	

individual's score is not already too heavily weighted one way or the other.

The teachers responded with variable degrees of satisfaction-dissatisfaction to their situations in general as indicated in Table X. The means of the teachers' responses to the total situation Column (Y_O), indicates that five teachers, numbers 7, 11, 15, 17, and 26 were more dissatisfied than satisfied with their situations. The aspects of the situation with which all members of this group showed more dissatisfaction than satisfaction included the students, the people of the community, and the school organization. One teacher indicated dissatisfaction with all nine aspects; the one showing least dissatisfaction was dissatisfied with six of the nine aspects.

Sixteen of the respondees indicated satisfaction with each of the nine aspects of the situation, while 24 who indicated satisfaction recorded dissatisfaction with one or more of the nine aspects of their situations.

As one examines these data it appears that teachers feel more satisfied with some aspects of the situations than others, and some teachers are more generally satisfied than others with the situation.

The means of the responses to the nine categories indicate that on the average more teachers were satisfied than dissatisfied with all nine categories. The rank order in which the teachers indicated most satisfaction with aspects of the situation are as follows: (1) administrative personnel and procedures, (2) faculty, (3) the particular position in which the teacher is, (4) the school organization, (5) the community in general, (6) instructional materials, (7) school plant, (8) the people of the community, and (9) the students. They were least satisfied with students.

A summary of each of the above nine categories follows. When percents are spoken of they were calculated by the process already indicated above.

1. The Administration

The analysis of the teachers' responses to individual items of the Inventory related to the administration (Table XI), shows that the teachers as a group indicated more dissatisfaction than satisfaction with no aspect of administration. Those items which were satisfactory to 50 percent of the teachers or more include the standing of the Board of Education in the community, the age, training, experience and philosophy of the superintendent. All other items ranged from 28 to 50 percent in satisfaction.

2. The Faculty

The analysis of the teachers' responses to the individual items of the Inventory related to the faculty of the school (Table XI) having a relatively high degree of satisfaction were the average age of the faculty, the number of Wisconsin Education Association memberships, and the availability of group insurance. A lower degree of satisfaction was registered for the person performing, the method used in performing, and the purpose and basis used in the evaluation of the work of the faculty. These scores in the aforenamed items ranged from 2 to 40 percent. In the same range of satisfaction were the number of faculty assistants available, and the type and competence of the help provided. Actual dissatisfaction was registered with regard to the adequacy of the professional library and the number of articles published by the members of the faculty.

3. The Particular Position

The analysis of the teachers' responses to individual items of the Inventory related to the particular position (Table XI), indicate a high degree of satisfaction with the degree of independence allowed and frequency of payment of salary these being given 82 and 74 percent satisfaction scores respectively. A general lower degree of satisfaction was registered for such aspects of the position as the size of the classes and the number of activities and the appropriateness of training to handle extra curricular activities. There was no item for which the number of teachers dissatisfied with the situation exceeded the number satisfied.

4. The School Organization

The analysis of the teachers' responses to individual items of the inventory related to the school organization (Table XI) shows that a high degree of satisfaction, ranging from 50 to 85 percent, was indicated for the hour of opening and closing school, length of class periods and the noon hour, and the vacation periods allowed. Within this range, the type of grading system used received the highest degree of satisfaction of all of the items on school organization. The importance of the grading system to the pupils, the aids available to teachers for the guidance of students, the value of participating in teacher meetings, and the adequacy of the school library in all areas received the lowest degrees of satisfaction, ranging from 1 to 5 percent. Other items of the Inventory in relation to the school organization that the teachers rated low in satisfaction were the rigid use of a fixed marking system, narrowness and inadequacy of course offerings to meet the student's needs, and the type and amount of guidance provided. Still others in this same category were the frequency, method of planning and the value of school assemblies. Low degrees of

TABLE XI

SELECTED ITEMS FROM THE TEACHFR SITUATION INVENTORY INDICATING THE PERCENT AND REASONS FOR SATISFACTION-DISSATISFACTION

SECTION I: THE COMMUNITY

Percent	Item	Reasons
	1. COMMUNITY LOCATION	
83.5	The State it is in:	Prefer to teach in home state. Like to live in warmer climate.
46.5	Section of the State it is in:	Like to be near large city. Prefer to be in Southern part of State.
	2. SIZE OF COMMUNITY	
36.5	The Population	Prefer to live in larger cities.
	3. COMMUNITY ACCESSIBILITY	
53.5	Adequacy, Condition of Highways	Good roads were generally available. Some need of road and street repair.
- 1.0	Convenience of Bus Schedules	Poor bus schedules.
-26.5	Convenience of Train Schedules	No convenient train schedules.
	4. ADJACENCY OF COMMUNITY	
49.0	Nearness to Shopping Centers	Near good shopping centers.
	5. GEOGRAPHY OF THE COMMUNITY	
39.0	Its Topography	Like the lakes, woods and hills.
	6. COMMUNITY APPEARANCE	
61.0	Its Cleanliness	Most towns and cities are kept very clean, some need to get funds to make improvements.
	7. INDUSTRIES OF THE COMMUNITY	
29.0	Their Contribution to Progress	Industry helps the cities. Need more interest by all people in community planning.

8. COMMUNITY INSTITUTIONS

65.5	The Diversity of Churches	Adequate to numerous.
58.0	Activity of the Churches	Very active to passive. Should coordinate activities of church with school.
63.5	Availability of Your Denomination	Not always available.

9. COMMUNITY HEALTH SERVICES

52.0	Availability of Good Doctors	Available for the most part, some communities need more doctors.
44.5	Availability of Good Dentists	Adequate to fair.

10. PUBLIC LIBRARY

29.0	Its Convenience of Opening Hours	No library, poor that are available, poor schedule, poor service.
16.5	Its Adequacy of Materials	Materials not adequate.

12. SOCIAL GROUPS, COMMUNITY CLUBS

36.5	Their Diversity	Many and varied; need more social groups for young people.
39.0	Their Purposes for Organization	Have social purposes by and large; need wider aims.
39.0	Their Availability to Teachers	Easily available; teachers can attend but not overly encouraged by the community.
40.0	Their Contribution to Community	Some have purposes; others purely social.

13. LIVING QUARTERS IN COMMUNITY

69.0	The Type You Found Available	Private homes, rooms, apartments.
81.0	The Suitability of its Location	Generally suitably located close to school.
75.5	The Modernness of Its Conveniences	Very modern.
76.0	The Restrictions, Regulations Asked	Few to no restrictions.
66.0	The Degree of Privacy Provided	All that can be desired; some need privacy.

14. MEAL ACCOMMODATIONS

The Convenience of Location

58.5 Good to excellent; walking distance from school.

17. CULTURAL OPPORTUNITIES IN COMMUNITY

Its Stage and Concert Offerings

-23.5 Extreme lack of concerts, plays. Need community interest to back the obtaining of it in the community.

Its Museums, Galleries, Zoos

-45.5 Non-existent.

Its Public Forums and Lectures

-27.5 None to moderate, offered by nearby institutions. Television coming in.

Its Provision for Adult Education

- 1.0 Agriculture, homemaking, vocational schools, more courses should be offered.

18. LEISURE TIME OPPORTUNITIES

Commercial Amusements Available

-19.0 General run of movies, bowling.

19. EXTRA JOB OPPORTUNITIES

Types of Extra Employment Available

-14.0 None to part-time work. More small businesses needed.

21. COMMUNITY DEMANDS ON TEACHERS

Community Responsibilities

35.0 Teacher left comparatively free, people recognize demands, teacher load.

Social Taboos, Bans on Behavior

43.0 Most anything within reason, very liberal.

Expectancies of Residence

30.0 Can leave for weekends, free to come and go.

SECTION II: PEOPLE OF THE COMMUNITY

2. LANGUAGE OF THE PEOPLE

The Quality of Everyday Usage

-32.0 Low level of English spoken generally, poor grammar usage. Need to change attitude of people toward language.

5. PEOPLE'S CULTURE

Their General Educational Level

- 5.5 Educational level is low.

Their Interests in the Arts

-22.0 Very little interest, people not interested in finer things of life. Need clubs, churches that would build this up.

7. CHARACTERISTICS OF THE PEOPLE

Their Hospitality

61.0 Extremely friendly generally; should make new teacher more welcome.

Their Friendliness

60.0 Good, congenial and helpful.

Their Cooperation with the School

29.0 Excellent to take school for granted.

Their Pride in Home and Community

60.0 Takes pride in immediate community; not the wide community.

SECTION III: SCHOOL PLANT

1. SITE OF SCHOOL PLANT

Its Convenience of Location

77.5 Most generally centrally located within walking distance.

Its Adequacy of Playground

25.0 Good to small, traffic hazards, more space generally needed.

Its Appearance of Grounds

49.0 Good generally, playgrounds need repair.

3. HEATING OF THE BUILDING

The Type in Use

50.0 Good, usually warm, well controlled.

4. LIGHTING IN THE BUILDING

The Type in Use

22.0 Mostly direct, single lights.

The Adequacy of Performance

19.5 Excellent to very poor. Halls need light.

The Convenience for Use

50.0 Good, to fair, to lack of proper outlets.

9. FACULTY ROOMS

The Number of Them

-43.5 No faculty room; need some room for faculty privacy. Felt the need of this very strongly.

The Degree of Privacy Afforded

-37.0 No privacy, boiler room only place to take a smoke and relax.

The Regulations for Their Use

-18.0 Not many regulations, can't smoke in some; poor furniture in them.

10. CUSTODIAL SERVICE

Its Adequacy

43.0 Excellent to fairly adequate, lax, uncooperative janitors. Should be more conscientious.

Its Competency

41.0 Very good to very little done, rooms not kept clean, too old men.

Its Cooperativeness with Faculty

44.5 Hold the efficient ones and get rid of the uncooperative ones.

SECTION IV: THE ADMINISTRATION

1. THE BOARD OF EDUCATION

Its Standing in the Community

60.5 Well liked, looked up to, respected, good.

Its Educational Philosophy

46.5 Conservative to progressive to let alone policy. Need people who understand problems.

3. THE SUPERINTENDENT

His Age

76.5 Mature in judgments; 35 to 60 years old.

His Years in Present Position

56.0 From 2 to 25 years.

His Years Training and Degrees

67.5 Sufficient, Masters to Ph. D.'s, mostly Master's degrees.

His Kind and Amount of Experience

74.5 Wide experience, range of 6 to 35 years experience.

His Educational Philosophy

71.0 Conservative to progressive; greater dignity in his actions. Broader philosophy needed.

6. ADMINISTRATIVE POLICIES

Standards for Student Promotion

30.5 No failing policy; level of scholarship should be improved; definite standards for promotion should be set and maintained.

Standards for Student Eligibility

32.5 Poor students passed so they can make the team; WIAA standards.

Teacher Experimentation, Research

36.5 Given considerable leeway; up to individual teacher so long as there is discipline; often encouraged.

Teacher Tenure, Seniority Rights

28.0 Fair to good.

Use of Democratic Procedures

50.0 Administrator makes final decisions; teachers encouraged to make suggestions; student's through council have part in decision making; should make democratic procedure sincere; lack of faculty planning.

SECTION V: SCHOOL ORGANIZATION

1. THE SCHOOL YEAR

The Vacations Allowed

62.5 Christmas and spring.

2. DAILY SESSIONS

The Hour of Opening

54.5 Range of 7:55 A.M. to 8:55 A.M.

The Length of Periods

55.5 Range of 43 to 60 minutes; prefer periods of varied length for different groups.

The Length of Noon Hours

56.0 Range of 30 minutes to 1-1/2 hours.

The Hour of Closing

61.0 Range between 3:30 P.M. to 4:30 P.M. Prefer 3:30 to 3:45.

3. SCHOOL GRADING SYSTEM

The Type Used

Letter grade A, B, E, D, F. More flexibility in grading systems.

The Rigidity of Its Use

25.5 Very rigid.

The Importance it Holds with Pupils

1.0 Pupils do not pay too much attention to it; know they won't fail. Other pupils put too much emphasis on grades.

4. SCHOOL CURRICULUM

The Broadness of Offerings

13.5 Very little to many choices possible. No languages offered. Commercial and shop courses needed.

Its Adequacy for Student Needs

13.5 Course offerings needed to meet the need in Agriculture, Shop, Home Economics.

Its Requirements and Electives

23.5 Not enough electives.

5. SCHOOL EXTRA-CURRICULUM

The Diversity, Variety of Offerings

49.0 Many activities offered such as Music, Speech, Athletics and many others.

The Adequacy for Student Needs

40.0 Does not take care of the needs of all students. Rural students not able to participate.

The Balance Among Activities

33.0 Too much emphasis on sports, Agriculture and Music. Need balance.

6. STUDENT GROUPING IN SCHOOL

The Bases Used to Group

22.5 I.Q. performance; grade level plus much guesswork. Need ability grouping.

The Purposes for Doing It

33.0 To make it easy to locate students; allow good students to progress.

7. STUDENT CONTROL

The Type Used in Whole School

28.5 Student council; purposes of student council not clearly defined. Establish policy on role of student council and furnish good guidance for it.

The Nature of Regulations Imposed

13.0 No set rules. Rules should be made by teachers, pupils and administration.

8. SCHOOL ASSEMBLIES

Their Frequency

16.0 Once or twice a month, not often enough, more planning needed by all concerned.

Their Purpose

28.5 For giving directions, instruction and recreation. Goals should be set with students in mind.

Their Value to the Students

18.0 Some excellent, others a waste of time.

The Responsibility for Organizing

15.0 Principal plans most, students and teachers others. Should be cooperative effort.

9. SCHOOL GUIDANCE OF STUDENTS

The Type and Amount Provided

12.0 Need of guidance director; mostly done by teachers.

The Aids Available to Teachers

1.0 Few to none provided. Need for material such as occupational information.

11. TEACHERS MEETINGS

Their Frequency

33.5 Formal and does not accomplish much. Two meetings a month. Need more well planned meetings.

Their Purpose

18.0 Discuss school problems, that could be taken care of in other ways. Teachers wont express themselves freely.

Their Procedure

18.0 Semi-democratic, more teacher participation needed.

The Value to Participants

3.5 One person gets the floor and keeps it. Those not directly concerned should be excused. Absolutely of no value.

SECTION VI: INSTRUCTIONAL MATERIALS

1. SCHOOL LIBRARY

Its Size and Number of References

19.0		Restricted facilities to very good facilities. Library improvement program should be set up.
3.5	Its Adequacy for all Areas	Limited in some fields, need source books in Literature and Social Sciences.
14.5	Its Recency of Books and Materials	Books are too old. Need fresh materials.
12.5	Its Variety and Range of Difficulty	Should have books for ranges of ability.
31.0	Its Suitability of Periodicals	Excellent to very limited selection. Need a variety in a central place.

2. AUDIO-VISUAL MATERIALS

56.5	Their Appropriateness	Service available, average use is made of it.
46.5	Their Condition, Ease of Operation	Excellent, average, good to never works right.
52.0	The Regulations for Use	Not strict, must sign schedule.
30.5	The Budget for Materials	Not enough money to unlimited.

3. TEXT MATERIALS

43.0	The Quantity Furnished	Not sufficient, limited; all new books.
52.5	Their Ownership	Furnished by school.
38.0	Their Appropriateness	Generally out of date and not enough range of difficulty.
35.5	Their Condition, Worn, etc.	Cheap, worn, to good to excellent.

4. CLASSROOM AIDS—LABORATORY

16.0	The Equipment Furnished	Could use more. From poor to fair to good and better than average.
11.5	Their Adequacy, Number and Size	From not enough, to average to excellent.
16.0	Their Condition	Poor, to good, to excellent.

5. STANDARDIZED TEST MATERIALS

24.5	The Extent of Use	Testing program in operation; left up to teachers.

The Purposes of Administration

23.5 For classification, diagnosis and placement; no one seems to know.

6. SUPPLEMENTARY MATERIALS

Provision for Field Trips

15.0 School buses available at times; field trips should be used more by teachers.

Provision for Outside Talent

15.0 Some through University of Wisconsin. Provision should be made for at least limited program.

SECTION VII: THE FACULTY

1. STRUCTURE OF FACULTY

Total Number of Individuals

44.5 Range of 5 to 62.

The Average Age

66.0 35 years; would prefer younger people to work with.

Their Age Range

58.0 Range of 21 to 60.

The Sex Distribution

42.0 50 - 50 to too many women.

Their Marital Status

29.5 Most men married, women single. Either more single men or none.

4. RESIDENCE OF FACULTY

The Percent Living Out of Town

59.5 20 to 25% live out of town. Should live in town and as part of community.

The Percent of Home Teachers

30.0 From none to 85%.

6. FACULTY PROFESSIONAL STATUS

The Number of WEA Memberships

64.5 Twenty-five teachers reported 100% in WEA.

The Number of Union Memberships

1.0 None.

The Number of Earned Recognitions

61.0 None to few, if any.

The Number Having Articles Published

- 6.0 None, few, 2 or 3.

7. ASSISTANTS TO FACULTY

The Number Available

2.5 None to one who is overworked. Need more faculty help. Have a time when faculty could call on help.

10.0	The Type of Help Provided	Student help.
12.5	The Competence of the Help	Poor to very competent, very willing to help.

8. SERVICES TO FACULTY

64.5	Availability of Group Insurance	Largely Blue Cross.
65.5	Provisions for Sick Leave	Average 5 days.
56.5	Rules on Emergency Leave	Very reasonable and liberal and understanding.
- 1.0	Adequacy of Professional Library	From good, adequate to very poor. Need material which relates to school problems.

9. SELECTION OF FACULTY

53.5	The Method Used	Personal interview with superintendent.
52.5	The Bases for Choice	Qualifications, personality and recommendations.

10. EVALUATION OF WORK OF FACULTY

17.5	The Method of Evaluation Used	None that is known; principal visits classroom; informal; not informed how teachers were evaluated.
33.0	The Purpose for It	Merit system; improve quality of faculty; promotion and raises.
18.5	The Bases Used in Evaluating	Educational theory; experience, or impressions of principal. Prefer teacher would be informed of basis of evaluation. Valuations could be more specific to help the teacher.
36.5	The Person Doing It	Curricular worker, principal, superintendent, supervisor.

SECTION VIII: THE STUDENTS

1. STUDENT ENROLLMENT

45.5	The Total Number	85 to 1600.
30.5	The Majority Sex	About equal.
32.5	The Teacher-Pupil Ratio	One teacher to 25 to 35 pupils. Prefer smaller groups of 20 to 25.

2. STUDENT RESIDENCES

The Distance from School

-10.0 Some come as far as 40 miles, too long bus ride. Bus routes are too long.

The Condition of the Homes

13.0 Poor, good, very comfortable.

3. STUDENT PERFORMANCE

The Persistency of Attendance

- 4.5 Too many absences, need close check. Classes should be made more interesting so students would be eager to attend.

The Level of Scholarship

-10.0 Generally low, few real students, no apparent ambition, little self direction. Great need for conscientious study.

4. STUDENTS' TRAITS

Their Behavior, Manners

12.5 Well mannered to rude, uncouth. Students show lack of home training. Parents should teach children manners.

5. STUDENT ATTITUDES

Those Toward Fellow Students

45.5 Generally good to too many cliques. Little concern for other students' feelings. Could be improved with right leadership.

Those Toward Learning

-15.0 No desire to learn, attend because they have to. Teacher must motivate to arouse desire for learning.

Those Toward Faculty

22.5 Good, to more respect needed, teacher can't let down for a minute.

6. STUDENT CONTRIBUTIONS

Percent Going to College

8.0 5 percent to 30 percent. Need to stress importance of education.

SECTION IX: THE PARTICULAR POSITION

1. YOUR TEACHING LOAD

The Appropriateness to Your Training

54.5 Good, very good, excellent.

The Number of Classes

43.0 Average 5 per day.

The Size of the Classes

14.0 Most generally too large, range 14 to 40, prefer 20 to 25.

2. YOUR EXTRA-CURRICULAR LOAD

The Number of Activities

22.0 Average of two. Some teachers carry too heavy a load in this area.

28.0 The Appropriateness to Your Training
Wide training should be required to handle various activities. Prefer student council training, pep club.

3. YOUR ADMINISTRATIVE LOAD

26.5 Amount of Home Room Responsibilities
Home room, detention hour. Need to be organized and planned.

32.0 Number of Noon-After School Duties
Noon duty, supervise school social functions, playground. Such work is distributed among teachers. Would prefer to have it distributed fairly.

39.5 The Number of Records and Reports
Just grades, 4 to 5 reports of various kinds a month.

49.0 The Times Lesson Plans are Required
None, checked every week, recommended by supervisor. Prefer not to have to submit them in too much detail.

82.0 The Degree of Independence Allowed
Very much on your own.

4. YOUR INSTRUCTIONAL MATERIALS

25.0 The Appropriateness of Other Aids
Should be more suitable; more maps, references and periodicals needed. Less difficult material needed for students of less ability.

5. YOUR COMMUNITY DUTIES

26.0 Their Nature
None. Entirely voluntary, exhibits, talks P. T. A.

6. YOUR SALARY

24.0 The Amount
Generally ranged from $2400 to $3450. Would prefer from $2600 to $3600.

74.0 The Frequency of Payment
Generally once a month. Choice of 9 to 12 months. Prefer two times per month.

21.0 The Bases for Determining
No basis. Salary schedule. Experience and training. Prefer single schedule.

27.5 The Number of Deductions
Health insurance, retirement, income tax. Prefer less deduction for retirement.

7. YOUR CONTRACT PROVISIONS

39.0 Their Exact Nature
None to very exact. Extra curricular duties not specified. Prefer to have more stated exactly what one must do.

37.0 Their Justness
Generally very fair.

	8. YOUR PROMOTION	
15.0	The Form it Would Take, its Nature	Raise in salary.
37.5	The Basis for Granting	Experience, competent performance, extra-curricular work. Based too much on discipline.
27.0	The Amount of Salary Increments	Range of $100 to $325 a year. Prefer set salary schedule so all teachers will know.
	9. FURTHER REQUIREMENTS	
13.5	The Teaching Method Desired	Left up largely to the individual teacher. Principal desires teacher control students. Prefer more supervision. More pupil planning.
53.5	The Student Relations Desired	Friendly but keep proper professional distance.
31.5	The Type of Discipline Wanted	Up to the individual teacher. Prefer more administrative backing.
31.0	The Committee Work Involved	Voluntary basis. Many committees working continually on problems and research.

satisfaction were indicated for such items as the importance of the grading system to the pupils, occupational information available, the guidance of students, and the value of teacher meetings to the participants. There were no items of the school organization that received more dissatisfaction responses than satisfaction responses.

5. The Community

The analysis of the teachers' responses to individual items of the Inventory related to the community (Table XI) indicate some dissatisfaction with convenience of bus and train travel, cultural opportunity, leisure time opportunity, and extra job opportunities. Those aspects of the community toward which the teachers indicated a high degree of satisfaction (75 percent or higher) were the state in which they taught, accessibility of living quarters, moderness of living quarters and the paucity of restrictions in their living quarters.

6. The Instructional Materials

The analysis of the teachers' responses to individual items of the Inventory related to the instructional materials (Table XI) indicated a relatively low degree of satisfactions with the appropriateness of supplementary aids to teachers. They indicated that there was a need for more suitable aids such as maps, charts, references, and periodicals. There was also need for graded material of various reading difficulty so as to meet the needs of all the students more adequately, according to the teachers responding to the inventory.

7. The School Plant

The analysis of the teachers' responses to individual items of the Inventory related to the school plant (Table XI) shows some high degrees of dissatisfaction. Those items which received high dissatisfaction scores (ranging from 18.0 to -43.5 percent) were the lack of a faculty room in which to have some privacy, and the restrictions placed on the use of a faculty room where there was one. The items which received high satisfaction scores were the location of the school plant, the heating of the building, the convenience of light, and the appearance of the playground. Lower satisfaction scores were indicated for the inadequacy of the playground, the type of lighting used, insufficient light for halls and stairways, and the inadequacy, incompetency and uncooperativeness of custodial service. These lower satisfaction scores ranged from 20 to 45 percent.

8. The People of the Community

The analysis of the teachers' responses to the individual items of the Inventory related to the people of the community (Table XI) indicates dissatisfaction with the people of the community with respect to their language usage, educational level, and interest in the arts. The friendliness and hospitality of the people were given 61 percent and 60 percent satisfaction scores, respectively.

9. The Students

The analysis of the teachers' responses to individual items of the Inventory related to the students (Table XI) indicates a generally low degree of satisfaction or a moderately low dissatisfaction with the various aspects relative to students' performance, attendance, level of scholarship, attitude toward learning, and the distance of pupils from school as the latter affects the various aspects of the students' school life. These items had a range of -4.5 to -15.0 percent of dissatisfaction. The low satisfaction scores were indicated in the areas of personal traits and manners of the students, such as behavior and manners, attitudes toward faculty members and fellow students, and the conditions of the homes.

Analysis of the Responses of the Fifteen Most Satisfied and Dissatisfied Subjects

A study was also made of the relation between satisfaction-dissatisfaction scores and teacher efficiency as determined by supervisory ratings. Three criteria of efficiency were employed: (1) an Acceptability rating, (2) an M-Blank rating, and (3) a Composite of Various Measures. The satisfaction-dissatisfaction scores previously reported in Table X, were arranged in rank order (see Appendix B-2, B-3 and B-4 for the raw data). The most dissatisfied teacher, i.e., the teacher with the highest raw score, was given rank order one. This was done separately for each category. The least dissatisfied teacher, i.e., the teacher with the lowest raw score, was ranked forty-fifth. A summary of these scores is given in Tables XII, XIII, and XIV.

From examining the data in Tables XII, XIII, and XIV, it may be noted that no consistent pattern is revealed. The mean scores of the 15 most dissatisfied and the 15 least dissatisfied teachers in relation to the acceptability criterion showed that the 15 most dissatisfied teachers received a more favorable mean efficiency rating. In relation to the Composite (X_O) criterion the fifteen most dissatisfied teachers also received a more favorable mean efficiency rating. In rela-

TABLE XII

THE MEAN VALUE FOR THE 15 MOST DISSATISFIED AND THE 15 LEAST DISSATISFIED TEACHER SCORES CALCULATED FOR THE SEVERAL FACTORS AND THE CRITERION OF ACCEPTABILITY (X_3)

Environmental Factors		15 Most Dissatisfied		15 Least Dissatisfied	
		Mean Score	Acceptability Criterion	Mean Score	Acceptability Criterion
I.	Community	3.10*	2.20	1.83	2.27
II.	People of Community	3.19	2.13	1.90	2.20
III.	School Plant	3.45	2.33	1.64	2.40
IV.	Administration	2.87	2.13	1.31	2.33
V.	School Organization	3.07	2.54	1.63	2.13
VI.	Instructional Materials	3.26	2.20	1.62	2.13
VII.	Faculty	2.91	2.27	1.64	2.40
VIII.	Students	3.36	2.07	1.77	2.33
IX.	Particular Position	2.95	2.33	1.61	2.13
	Total Mean I - IX	3.13	2.24**	1.66	2.26

* This and all other values in this column represent mean scores (see table) for the evaluation assigned by the several teachers to each item.

**The population of the 15 highest and 15 lowest change from category to category; this is the difference in the mean criterion score.

TABLE XIII

THE MEAN VALUE FOR THE 15 MOST DISSATISFIED AND THE 15 LEAST DISSATISFIED TEACHER SCORES CALCULATED FOR THE SEVERAL FACTORS AND THE M-BLANK CRITERION (X_2)

Environmental Factors	15 Most Dissatisfied		15 Least Dissatisfied	
	Mean Score	M-Blank Criterion	Mean Score	M-Blank Criterion
I. Community	3.10	2.27	1.83	2.20
II. People of Community	3.19	2.33	1.90	2.27
III. School Plant	3.45	2.20	1.74	2.27
IV. Administration	2.87	2.47	1.31	2.20
V. School Organization	3.07	2.47	1.63	2.07
VI. Instructional Materials	3.26	2.53	1.62	2.07
VII. Faculty	2.91	2.47	1.64	2.07
VIII. Students	3.36	2.33	1.77	2.27
IX. Particular Position	2.95	2.53	1.61	2.48
Total Mean I - IX	3.13	2.40	1.66	2.21

TABLE XIV

THE MEAN VALUE FOR THE 15 MOST DISSATISFIED AND THE 15 LEAST DISSATISFIED TEACHER SCORES CALCULATED FOR THE SEVERAL FACTORS AND THE COMPOSITE CRITERION (X_0)

Environmental Factors	15 Most Dissatisfied		15 Least Dissatisfied	
	Mean Score	Composite Criterion	Mean Score	Composite Criterion
I. Community	3.10	2.45	1.83	2.48
II. People of Community	3.19	2.60	1.90	2.71
III. School Plant	3.45	2.48	1.64	2.64
IV. Administration	2.87	2.63	1.31	2.61
V. School Organization	3.07	2.73	1.63	2.62
VI. Instructional Materials	3.26	2.53	1.62	2.53
VII. Faculty	2.91	2.64	1.63	2.86
VIII. Students	3.36	2.58	1.77	2.61
IX. Particular Position	2.95	2.72	1.61	2.48
Total Mean I - IX	3.13	2.60	1.66	2.62

tion to the M-Blank criterion the fifteen most dissatisfied teachers received a less favorable efficiency rating. It should be noted that there is much inconsistency among the various individuals that constitute the two groups.

One would ordinarily assume that inasmuch as the conditions, by and large, in the smaller schools are not as satisfactory as in larger schools (certainly with respect to the physical plant materials of instruction and in some instances the character of the pupils) that teachers in smaller schools would show more dissatisfaction than teachers in larger schools. This suggestion is borne out by the data in Table XIVa. In general, the teachers almost without exception in the smaller communities are more dissatisfied than teachers in the larger communities. Approximately half of these teachers (7 out of 15) received higher efficiency ratings than the teachers in larger communities.

This is in accordance with expectation inasmuch as there are certain things with respect to which good teachers showed dissatisfaction as well as poor teachers.

Correlational Analysis

A further analysis of the data was made through use of correlational techniques. Correlations were calculated between the categories of Table XV and teacher efficiency. These range from -.06 to +.41; a number were near zero. From these data it appears that there is a positive relationship between teacher efficiency and the sort of people that make up the community. There is also a positive correlation between criterion X_1, which is the State Department rating, and between all aspects of the situation and criterion X_4 which was Montross' criterion. The highest positive correlations are between the School Organization Y_5, Instructional Materials Y_6, The Faculty Y_7, Mean Total Y_0, and X_1, and X_4. The low and zero correlations indicate no relationship. It should be noted that some of the relationships are negative, e.g., — indicating that high dissatisfaction in some instances is associated with efficiency.

Correlations between the M-Blank ratings and certain categories were among the highest. This supports one of the earlier findings wherein the 15 most dissatisfied teachers received less favorable M-Blank ratings than the 15 least dissatisfied. To gain a better understanding of efficiency between the M-Blank and responses of all the subjects to the four categories in which positive correlations were found, the multiple correlation technique was applied and the results are given in Table XVa.

It may be noted in this connection, that relatively high positive intercorrelations (ranging between .59 and .83) were found between the categories pertaining to the students and school organization, students and instructional materials, students and faculty, faculty and school organization, and instructional materials and school organization. These high intercorrelations would appear to indicate that teachers tended to respond to the various categories listed with a generalized attitude of satisfaction or dissatisfaction, and that this generalized attitude was reflected in efficiency ratings received on the M-Blank to a certain extent. The multiple correlation obtained between the M-Blank ratings and the responses to the four categories was 0.43.

Analysis of Responses of the Interviewed Sample

A further study was made of the fifteen subjects who had not returned the inventory. They were interviewed to obtain their reactions to selected aspects of their situations.

After rapport had been established they were asked the following questions:

1. With what aspects of the community would you approve or disapprove?
2. What aspects of the people of the community would you approve or disapprove?
3. What aspects of the school plant would you approve or disapprove?
4. With what aspects of the administration would you approve or disapprove?
5. With what aspects of the school organization would you approve or disapprove?
6. With what aspects of the instructional materials would you approve or disapprove?
7. With what aspects of the faculty would you approve or disapprove?
8. With what aspects of the student body would you approve or disapprove?
9. With what aspects of the particular position would you approve or disapprove?

For those items in each category to which the subject responded where satisfaction was expressed, a value of plus 1 was assigned. For those items to which the subject responded with dissatisfaction, a value of minus 1 was assigned. When the subject neither approved nor disapproved, a zero value was assigned. Satisfaction-dissatisfaction scores were computed for this three-point scale and the responses converted into a percent of satisfaction-dissatisfaction.

The individual responses to the selected items of the Inventory are given in Appendix B-5. A summary of the responses of the fifteen subjects are given in Table XVI. This table shows the sum of the approval, neutral, and disapproval responses for each category, and the mean percent of satisfaction-dissatisfaction for the fifteen subjects' response to this category.

The rank order in which the sample of fifteen

TABLE XIVa

THE MEAN VALUE FOR THE 15 SMALLEST SCHOOL COMMUNITIES AND THE 15 LARGEST SCHOOL COMMUNITIES CALCULATED FOR THE COMPOSITE CRITERION (X_0) AND THE SATISFACTION-DISSATISFACTION SCORES

Fifteen Smallest School Communities			Fifteen Largest School Communities		
Subject Number	Dissatisfaction-Satisfaction Scores	Composite Efficiency Rating (X_0)	Subject Number	Dissatisfaction-Satisfaction Scores	Composite Efficiency Rating (X_0)
37	2.70	4.00	33	1.84	2.50
18	2.94	2.75	38	2.55	3.50
25	2.39	3.00	5	2.53	1.00
17	3.06	1.75	6	1.79	2.00
9	2.04	2.00	13	1.53	2.00
42	2.52	3.25	44	2.21	2.00
26	3.18	2.00	41	2.39	1.65
24	2.93	2.00	23	1.94	3.25
27	1.83	2.50	10	2.41	4.67
2	2.11	3.75	3	2.55	2.00
22	2.99	2.50	39	2.09	2.50
29	2.59	2.75	28	1.95	3.50
16	2.85	2.50	35	2.62	2.00
15	3.42	3.00	4	1.67	2.25
19	2.96	2.75	14	1.68	3.00
Total	40.51	40.50		31.75	37.82
Mean	2.70	2.70		2.11	2.52

TABLE XV

MEASURES OF SATISFACTION-DISSATISFACTION CORRELATED WITH MEASURES OF TEACHING EFFICIENCY

N	Criterion		Com-munity Y_1	People of Com-munity Y_2	School Plant Y_3	Adminis-tration Y_4	School Organi-zation Y_5	Instruc-tional Materials Y_6	Faculty Y_7	Students Y_8	Partic-ular Position Y_9	Mean Total Y_0
45	Composite	(X_0)	.06	.03	.04	.04	.02	.01	-.06	-.03	.17	.05
45	Principals' Acceptability	(X_3)	.01	.13	.05	-.01	.04	.04	.00	-.01	.07	.03
45	Wisconsin Adaptation of the M-Blank	(X_2)	.02	.11	.13	.10	.27	.32	.29	.20	-.02	.26
36	Criterion	(X_1)	.16	.10	.08	.06	.01	.08	.04	.09	.13	.10
24	Criterion	(X_4)	.21	.17	.28	.20	.19	.22	.19	.08	.41	.26

TABLE XVa

CORRELATIONAL MATRIX FOR SELECTED CATEGORIES OF THE INVENTORY

		X_1	X_2	X_3	X_4	X_5
M-Blank	(X)	1.00	.27	.32	.29	.20
School Organization	(X)	.27	1.00	.83	.75	.59
Instructional Materials	(X)	.32	.83	1.00	.10	.67
Faculty	(X)	.29	.75	.10	1.00	.65
Students	(X)	.20	.59	.67	.65	1.00

Note: The multiple regression equation, when calculated, became

$$Z_1 = .61\ Z_2 + .06\ Z_3 + (-.21) Z_4 + .25\ Z_5$$

responded to the nine categories and the rank order in which the original group of forty-five responded are presented in Table XVII.

From inspection of the table it may be noted that the rank order of the mean responses of the fifteen subjects interviewed was identical with the original forty-five with respect to five categories: The Students, The People of the Community, The Community, The Particular Situation, and The Administration. Rank orders differed by 1 or 2 with respect to Instructional Materials, the Faculty, and the School Organization. A wide difference was found, however, with respect to the two groups' responses to the school plant. In general, the responses of those interviewed did not differ markedly from those filling out the Inventory. Those interviewed, however, rated all categories of the environment less satisfactory than did the original group.

Analysis of Case Studies

Ten case studies are here presented, which include the five most satisfied teachers and the five most dissatisfied teachers. One dissatisfied teacher was omitted because of the insufficiency of data available.

Each case study is presented in two parts. The first part presents general information relative to such items as age, sex, marital status, family data, parents' education, parents' occupation, teacher preparation, number of different schools attended, degree of self support, jobs held for pay, number of trips taken, veteran, and experience in youth leadership. The second part summarizes each subject's responses to the various categories of the Inventory (Appendix B-1). In this part of the case study, the major sources of satisfaction or dissatisfaction are identified, the mean efficiency rating is reported, and the relationship between satisfaction and efficiency rating indicated.

SUBJECT NUMBER 21

The analysis of the subject's responses to the various items of the Inventory (Table X) shows that satisfaction was expressed with such frequency and with such degrees of intensity as to indicate a generally positive response. The mean total percent of satisfaction is 76.5. In this sense this teacher is the most satisfied of the group of teachers studied in this investigation. The four categories to which the subject responded with the most satisfaction were the administration, 97.0 percent; the school plant, 97.0 percent; the instructional materials, 95.5 percent; and the school organization, 89.5 percent. In regard to the administration (Appendix B-1) the subject responded in a highly favorable degree to all items, including the board of education, the accreditation of the school, and the superintendent. Satisfaction was also registered for standards of student promotion and use of democratic procedures. The items of the school plant to which the subject responded in a highly favorable degree were the site of the school plant; the school building and floors; hearing; lighting; rooms in the building; toilet facilities; furniture; faculty rooms; and the adequacy, competency and cooperativeness of custodial service. There were no items of the school plant to which the subject responded in any degree of dissatisfaction.

This subject received a composite efficiency rating (X_O) of 2.75, which is slightly above average. This teacher ranked 28.5 from the top in the group of forty-five teachers.

SUBJECT NUMBER 13

The analysis of the subject's responses to the various items of the Inventory (Table X) shows that satisfaction was expressed with such frequency and with such degrees of intensity as to indicate a generally positive response. The mean total percent of satisfaction is 73.5. The five categories to which the subject responded with the most satisfaction were: the instructional materials, 92.0 percent; the students, 91.0 percent; the particular position, 91.0 percent; the school organization, 83.5 percent, and the faculty, 78.5 percent. The items in regard to the instructional material (Appendix B-1) to which the subject responded in a highly favorable degree were, all items in regard to the school library, audio-visual materials, text materials, aids for classroom and laboratory, and standardized test materials. Satisfaction was also registered with reference to provisions for field trips and outside talent. There were no items of the instructional materials to which the subject responded with dissatisfaction. The items in regard to the students to which the subject responded in a highly favorable degree were student enrollment, student residences and the performance of students in respect to attendance, level of scholarship, extra-curricular participation, extent of community activity, extent of social activities and percent employed. Satisfaction was also registered with reference to behavior, loyalty, school pride, cooperation and friendliness, as was satisfaction registered for student attitudes toward fellow students, faculty, and rival schools. The percent going to college was also given a favorable response. There were no items with reference to students to which the subject responded with dissatisfaction.

The subject received a composite efficiency rating (X_O) of 2.00 which places him 12th in the whole group of forty-five.

SUBJECT NUMBER 21

AGE	SEX	MARITAL STATUS	FAMILY DATA	PARENTS' EDUCATION	PARENTS' OCCUPATION
23	Female	Single	One older brother	Mother: completed twelfth grade Father: completed college	Father: Professor, University

TEACHER PREPARATION	NUMBER DIFFERENT SCHOOLS ATTENDED	DEGREE OF SELF-SUPPORT	JOBS HELD FOR PAY
Major: Home Economics Minor: Science	Elementary: 1 Secondary: 1 College: 1	Since College: 98%	Domestic: 2

NUMBER TRIPS TAKEN (NON-MILITARY)	VETERAN	EXPERIENCE OF YOUTH LEADERSHIP
Over 500 miles: 6 Foreign country: 1		Class office held: 1

SUBJECT NUMBER 13

AGE	SEX	MARITAL STATUS	FAMILY DATA	PARENTS' EDUCATION	PARENTS' OCCUPATION
25	Male	Single	One younger brother	Mother: completed normal school Father: completed eighth grade	Father: Foreman, paper mill

TEACHER PREPARATION	NUMBER DIFFERENT SCHOOLS ATTENDED		DEGREE OF SELF-SUPPORT	JOBS HELD FOR PAY	
Major: Chemistry Minors: Mathematics, General Science	Elementary:	1	College: 50%	Professional:	1
	Secondary:	2	Since College: 100%	Agricultural:	1
	College	1		Factory	1
				Manual	1

NUMBER TRIPS TAKEN (NON-MILITARY)	VETERAN	EXPERIENCE OF YOUTH LEADERSHIP	
None	18 months domestic service 12 months foreign service Navy Rank: Ph M 1/C	Camp counseling:	9 months
		Scout Leader:	7 years

TABLE XVI

SUMMARY OF THE RESPONSES OF THE FIFTEEN SUBJECTS OF THE INTERVIEW

Situations	P	%	O	N	%
I. The Community					
With what aspects of the Community would you approve or disapprove?					
a) State	12	80	0	3	20
b) Accessibility	6	40	0	9	60
c) Cleanliness	8	53	0	7	47
d) Church groups	9	60	6	0	0
e) Library	4	27	0	11	73
f) Health services	8	53	0	7	47
g) Social groups	7	47	0	8	53
h) Living quarters	9	60	0	6	40
i) Cultural opportunities	4	27	0	11	73
Total	67	440	6	62	413
Mean		52			48
II. People of the Community					
With what aspects of the people of the Community would you approve or disapprove?					
a) Quality of everyday language usage	2	13	0	13	87
b) Type of occupations	4	30	2	9	70
c) Average family income	4	36	4	7	64
d) Educational level	4	27	0	11	73
e) Interest in the arts	2	22	6	7	78
f) Behavior characteristics	3	43	8	4	57
g) Hospitality	8	57	1	6	43
h) Cooperation with school	7	58	3	5	42
Total	34	286	24	62	514
Mean		35			65
III. School Plant					
With what aspects of the School Plant would you approve or disapprove?					
a) Convenience of location	12	86	1	2	14
b) Freedom from hazards	8	65	2	5	15
c) Adequacy of heating	13	87	0	2	13
d) Adequacy of lighting	12	80	0	3	20
e) Adequacy of rooms	9	60	0	6	40
f) Adequacy of furniture	5	33	0	10	67
g) Provision for faculty privacy	4	27	0	11	73
h) Adequacy of custodial service	8	53	0	7	47
Total	71	491	3	46	289
Mean		61			39

P = Approval O = Neutral N = Disapproval

TABLE XVI (Continued)

Situations	P	%	O	N	%
IV. The Administration					
With what aspects of the Administration would you approve or disapprove?					
a) Education of the Board	9	75	3	3	25
b) Philosophy of the Board	10	83	3	2	17
c) Rules and regulations of the Board	13	87	0	2	13
d) Superintendent training and degrees	9	75	3	3	25
e) Superintendent hired and amount of experience	11	85	2	2	15
f) Superintendent educational philosophy	13	93	1	1	7
g) Administrative policy in regard to:					
1. Student promotion	5	22	0	10	67
2. Teacher experimentation	7	58	3	5	42
3. Use of democratic procedures	10	67	0	5	33
Total	87	656	15	33	244
Mean		73			27
V. School Organization					
With what aspects of the School Organization would you approve or disapprove?					
a) Type of grading system used	11	73	0	4	27
b) Rigidity of use of grading system	7	47	0	8	53
c) Broadness of curricular offerings	5	33	0	10	67
d) School grouping of students' purpose	5	50	5	5	33
e) Type of student control used	6	40	0	9	60
f) School assemblies, frequency and purpose	6	40	0	9	60
g) Type and amount of guidance provided	4	27	0	11	73
h) Guidance aids available to teachers	2	13	0	13	87
i) Health examinations	7	47	0	8	53
j) Frequency and purpose of teachers' meetings	5	33	0	10	67
k) Value of meetings to teachers	4	27	0	11	73
Total	62	430	5	98	653
Mean		39			61

TABLE XVI (Continued)

Situations	P	%	O	N	%
VI. Instructional Materials					
With what aspects of the Instructional Materials would you approve or disapprove?					
a) Library's size and number of references	3	20	0	12	80
b) Library's adequacy in all areas	2	13	0	13	87
c) Visual aids available	9	60	0	6	40
d) Quantity of textbooks	11	73	0	4	27
e) Adequacy of laboratory material	5	33	0	10	67
f) Extent of the use of standard test materials	4	27	0	11	73
g) Provision for field trips	7	47	0	8	53
h) Provision for outside talent	3	20	0	12	80
Total	49	293	0	76	507
Mean		37			63
VII. The Faculty					
With what aspects of the Faculty would you approve or disapprove?					
a) Age range	9	60	3	3	20
b) Training in years	10	67	0	5	33
c) Faculty experience in years	9	60	1	5	33
d) Percent of faculty living out of town	8	53	0	7	47
e) Amount of help for faculty use	3	20	0	12	80
f) Sick leave provisions	11	73	0	4	27
g) Adequacy of professional library	2	13	2	11	73
h) Method of choosing faculty	9	60	0	6	40
i) Method of evaluating faculty	5	33	5	5	33
j) Base used in evaluation	4	27	6	5	33
k) Purpose for evaluation	4	27	5	6	40
l) Friendliness of faculty	12	80	0	3	20
Total	86	573	22	72	479
Mean		54			46

TABLE XVI (Continued)

Situations	P	%	O	N	%
VIII. The Students					
With what aspects of the Student Body would you approve or disapprove?					
a) Pupil teacher ratio	6	40	0	11	73
b) Distance of students from school	5	33	0	10	67
c) Student performance as to:					
1. Attendance	2	13	0	13	87
2. Level of scholarship	3	20	0	12	80
3. Behavior	3	20	0	12	80
4. Cooperation	6	40	0	9	60
5. Appearance	5	33	0	10	67
6. Friendliness	11	73	0	4	27
d) Student attitudes in regard to:					
1. Toward fellow students	7	47	0	8	53
2. Toward learning	2	13	0	13	87
3. Toward faculty	5	33	0	10	67
4. Towards rival schools	8	60	0	7	47
Total	63	425	0	119	795
Mean		35			65
IX. The Particular Position					
With what features of the Particular Position would you approve or disapprove?					
a) Number of classes	8	53	0	7	47
b) Appropriateness of your training	12	80	0	3	20
c) Size of classes	5	33	0	10	67
d) Time spent on extra-curricular activities	3	20	2	10	67
e) Home room responsibility	4	27	7	4	27
f) Requirement of lesson plans	8	53	3	4	27
g) Time consumed by community duties	8	53	0	7	47
h) Bases for determining salary	5	33	6	4	27
i) Nature of contract provisions	10	67	3	2	13
j) Bases for granting promotion	6	40	4	5	33
k) Teaching method desired	8	53	5	2	13
l) Student relations desired	12	80	2	1	7
m) Type of discipline wanted	7	47	0	8	53
n) Committee work required	8	53	2	5	33
Total	104	692	34	72	481
Mean	59				41

TABLE XVII

RANK ORDER OF THE RESPONSES OF THE FIFTEEN SAMPLE AND ORIGINAL FORTY-FIVE SUBJECTS TO SELECTED AND ALL THE CATEGORIES OF THE INVENTORY RESPECTIVELY

Category	Rank Order of Interviewed Sample of 15	Rank Order of Original Sample of 45
Administration	1	1
School Plant	2	7
Particular Situation	3	3
Faculty	4	2
Community	5	5
School Organization	6	4
Instructional Materials	7	6
People of the Community	8.5	8
The Students	8.5	9

SUBJECT NUMBER 32

The analysis of the subject's responses to the various items of the Inventory (Table X) shows that satisfaction was expressed with sufficient frequency and with such degrees of intensity as to indicate a generally positive response. The mean total percent of satisfaction is 72.0. The five categories in rank order to which the subject responded with the most satisfaction were instructional materials, 91.5 percent; school plant, 89.0 percent; students, 88.0 percent; school organization, 86.0 percent; and administration, 78.0 percent. The items in regard to the instructional materials (Appendix B-1) to which the subject responded in a highly favorable degree were all items in regard to the school library, audio-visual material, text material, standardized test material and supplementary material. In addition, in regard to the instructional materials, degrees of satisfaction were registered for equipment furnished, condition and appropriateness of aids to classrooms and laboratories. The subject responded with dissatisfaction to only one item in reference to instructional materials which was, the amount, number and size of aids for classroom and laboratories. The items in regard to the school plant (Appendix B-1) to which the subject responded in a highly favorable degree were all items in regard to the site of school plant, the school building and floors, heating, lighting, rooms in the building, toilet facilities, furniture, faculty rooms and custodial service. There were no items of the school plant to which the subject responded in degree of dissatisfaction.

The subject received a composite efficiency rating (X_O) of 3.00, which places him 32.5 in the whole group of forty-five.

SUBJECT NUMBER 4

The analysis of the subject's responses to the various items of the Inventory (Table X) shows that satisfaction was expressed with sufficient frequency and with such degrees of intensity as to indicate a generally positive response. The total mean percent of satisfaction was 66.5. The five highest satisfaction scores were with reference to students which were 100 percent, the school plant which was 83.0 percent, the community which was 78.0 percent, the people of the community which was 76.0 percent, and the administration which was 75.0 percent.

There were no items with regard to the students to which this teacher responded with dissatisfaction.

This subject responded with dissatisfaction to only one item with reference to the school plant (Appendix B-1), namely, the cooperativeness of the custodial service. All other items were scored in such a manner as to indicate a high degree of satisfaction.

This subject received a composite (X_O) teacher efficiency rating of 2.25 which places him 18th in the whole group of forty-five.

SUBJECT NUMBER 14

The analysis of the subject's responses to the various items of the Inventory (Table X) shows that satisfaction was expressed with such frequency and with such degrees of intensity as to indicate a generally positive response. The total mean percent of satisfaction was 64.0. The five highest scores of satisfaction in rank order are: the administration, 100 percent; school organization, 78.5 percent; faculty, 75.5 percent; instructional materials, 74.0 percent, and the students, 71.5 percent.

The subject responded to all the items of the administration (Appendix B-1) with a 100 percent of satisfaction.

The items to which the subject responded with satisfaction in regard to the school organization were: the school year, daily sessions, grading system used, school curriculum, the school extra-curricular, methods of student control, and school assemblies. In addition, satisfaction was expressed with the aids available to teachers for school guidance of students, pupil services in relation to health, nutrition and employment aids, also teacher meetings as to frequency, purpose, procedure and time of meeting, along with public programs given by the school.

The items to which the subject reponded to as being dissatisfactory were: the type and amount of guidance provided, and the person or persons responsible for the guidance of students. In addition dissatisfaction was expressed as to the value to the participant of teachers' meetings.

This subject received a composite (X_O) teacher efficiency rating of 3.00, which places him 32.5 in the whole group of forty-five.

SUBJECT NUMBER 15

The analysis of the subject's responses to the various items of the Inventory (Table X) shows that dissatisfaction was expressed with such frequency and with such degrees of intensity as to indicate a generally negative response. The mean total percent of dissatisfaction was -21.0. The five categories to which the subject responded with the most dissatisfaction were: the instructional materials, -66.0 percent; the students, -54.0 percent; school plant, -45.0 percent; administration, -44.5 percent; and the school organization, -27.0 percent.

The items in regard to the instructional materials (Appendix B-1) to which the subject responded in a dissatisfactory manner were all

SUBJECT NUMBER 32

AGE	SEX	MARITAL STATUS	FAMILY DATA	PARENTS' EDUCATION	PARENTS OCCUPATION
22	Female	Single		Mother: completed 2 years college Father: completed eighth grade	Father: custodian local school Mother: teacher

TEACHER PREPARATION	NUMBER DIFFERENT SCHOOLS ATTENDED	DEGREE OF SELF-SUPPORT	JOBS HELD FOR PAY
Major: Business Education Minor: Economics English	Elementary: 1 Secondary: 2 College: 1	Since college: 100%	Professional: 1 Clerical: 1

NUMBER TRIPS TAKEN (NON-MILITARY)	VETERAN	EXPERIENCE OF YOUTH LEADERSHIP
None		Sunday school teaching: 2 years Class office: 3 years Pres., Young Peoples Christian Endeavor: 1 year

SUBJECT NUMBER 4

AGE	SEX	MARITAL STATUS	FAMILY DATA	PARENTS' EDUCATION	PARENTS' OCCUPATION
25	Male	Single	Three older sisters	Mother: completed eighth grade Father: completed twelfth grade	Mother: inspector Parker Pen Co. Father: retired railroad clerk

TEACHER PREPARATION	NUMBER DIFFERENT SCHOOLS ATTENDED	DEGREE OF SELF-SUPPORT	JOBS HELD FOR PAY
Major: Physical Education Minor: History Science	Elementary: 2 Secondary: 1 College: 1	High school: 10% College: 90% Since college: 100%	Professional 1 Proprietary 1 Sales work 2 Domestic 1 Manual 3

NUMBER TRIPS TAKEN (NON-MILITARY)	VETERAN	EXPERIENCE OF YOUTH LEADERSHIP
Over 500 miles: 2	Domestic service: 18 months Foreign service: 8 months Branch: Army Air Force Rank: Corporal	Tutoring others: 1-1/2 yrs. Sunday school teaching: 2 yrs. Camp counseling: 3 months Class offices held: 1 year Scout leader 1 year Committee member: 1-1/2 yrs. Guidance director: 4 months

SUBJECT NUMBER 14

AGE	SEX	MARITAL STATUS	FAMILY DATA	PARENTS' EDUCATION	PARENTS' OCCUPATION
24	Male	Married	Two younger brothers Two younger sisters	Mother: completed business college Father: completed college and seminary	Father: Minister

TEACHER PREPARATION	NUMBER DIFFERENT SCHOOLS ATTENDED	DEGREE OF SELF-SUPPORT	JOBS HELD FOR PAY
Major: Business Education Minor: Economics	Elementary: 2 Secondary: 1 College: 1 Trade School: 1	High School: 100% College: 100% Since College: 100%	Professional 2 Proprietary 1 Clerical 1 Factory 1 Manual 1

NUMBER TRIPS TAKEN (NON-MILITARY)	VETERAN	EXPERIENCE OF YOUTH LEADERSHIP
Over 500 miles: 10 Foreign country: 1	Domestic Service: 20 months Branch: Army Air Force Rank: PFC	Sunday school teaching: 1 year Class offices held: 2 years President of religious groups at University: 1 year

SUBJECT NUMBER 15

AGE	SEX	MARITAL STATUS	FAMILY DATA	PARENTS' EDUCATION	PARENTS' OCCUPATION
26	Male	Married	One older brother	Mother: completed seventh grade Father: completed third grade	Father: Farmer Mother: Domestic worker

TEACHER PREPARATION	NUMBER DIFFERENT SCHOOLS ATTENDED	DEGREE OF SELF-SUPPORT	JOBS HELD FOR PAY
Major: American Institutions Minor: English Economics	Elementary: 3 Secondary: 2 College: 1 Trade school: 1	High school: 25% College: 100% Since college: 100%	Professional 1 Proprietary 1 Saleswork 1 Agricultural 1 Factory 2 Manual 1

NUMBER TRIPS TAKEN (NON-MILITARY)	VETERAN	EXPERIENCE OF YOUTH LEADERSHIP

items in regard to the following: school library, audio-visual materials, text materials, and laboratory classroom aids. The subject responded with satisfaction to only one item in regard to the instructional materials and it was the provision for transportation for field trips.

The items to which the subject responded to as dissatisfactory in regard to the students were: the student's performance as to attendance, level of scholarship, extra-curricular participation, students' behavior, school pride, their cooperation, appearance and friendliness, attitude of students toward fellow students, learning, the faculty, and rival schools. Further dissatisfaction was expressed with the percent going to college, in National Student Association and the number of inter-scholastic championships. There were no items of the students to which the subject responded to in any degree of satisfaction.

The subject received a composite efficiency rating (X_O) of 3.00, which places him 32.5 in the whole group of forty-five.

SUBJECT NUMBER 7

The analysis of the subject's responses to the various items of the Inventory (Table X) shows that dissatisfaction was expressed with sufficient frequency and with such degrees of intensity as to indicate a generally negative response. The mean total percent of dissatisfaction was -13.5 percent. The five categories in which the subject indicated the most dissatisfaction were: the school plant, -26.5 percent; the students, -19.0 percent; administration, -16.5 percent; the particular position, -14.0 percent; and the school organization, -13.0 percent.

Items of the school plant (Appendix B-1) to which the subject responded to as dissatisfactory were: degree of moderness of the school building, provisions for ventilation, inadequate heating, size of the rooms and their decoration, arrangement of the room and freedom from disturbances, usefulness of the furniture, number of faculty rooms with provisions of privacy, and the cooperativeness and competency of the custodial service. There were no items of the school plant to which the subject responded with a degree of satisfaction.

The individual items, in regard to the students, to which the subject responded with a degree of dissatisfaction were: the performance of the students in regard to attendance and level of scholarship, extent of social activities, behavior of students, students' cooperation and appearance, and attitude of students toward learning and faculty. There were no items of the students to which the subject responded in any degree of satisfaction.

The subject received a composite efficiency rating (X_O) of 3.50, which places him 39th in the whole group of forty-five.

SUBJECT NUMBER 26

The analysis of the subject's responses to the various items of the Inventory (Table X) shows that dissatisfaction was expressed with sufficient frequency and with such degrees of intensity as to indicate a generally negative response. The total mean percent of dissatisfaction was -9.0. Five scores of dissatisfaction in rank order were: the community, -25.5 percent; school plant, -20.5 percent; people of the community, -14.5 percent; school organization, -14.0 percent; and the students, -6.5 percent.

The items with reference to the community with which the subject responded to a degree of dissatisfaction were the section of the state in which the community was located, the size of the population of the community along with the community area and spread. Dissatisfaction was also registered with the community accessibility, the surrounding area of the community and the type of industries found. This subject was also dissatisfied with the activities of the churches, schools other than public, along with the availability of good dentists and doctors, and the convenience of a hospital. Still this subject was dissatisfied with the public library as to convenience of opening hours, adequacy of materials and the recency of the books, and other materials. Furthermore, dissatisfaction was registered for social groups as to their diversity, purpose and contribution to the community. The lack of privacy in living quarters, the meal accommodations as to type found available, and the quality of food was indicated as dissatisfactory by this subject. As to rentals, prices paid for meals and groceries, and the cost of sundries this subject registered dissatisfaction. Finally, the subject registered dissatisfaction with the cultural opportunities in the community, the distinctions of the community, leisure time opportunities in regard to amusements and recreation available, along with extra job opportunities.

The items with reference to the community with which this subject expressed degrees of satisfaction were: the state in which the community was located; the type, location, modernness, restrictions, and the degree of privacy provided as to living quarters. As to location of meal accommodations it was found to be satisfactory along with the community demands made on teachers in regard to community responsibility, social taboos, bans on behavior, and other items of this nature.

The items with reference to the school plant (Appendix B-1) with which this subject registered degrees of dissatisfaction were, the type of construction of the school building, along with the rooms. In addition, the subject registered

SUBJECT NUMBER 7

AGE	SEX	MARITAL STATUS	FAMILY DATA	PARENTS' EDUCATION	PARENTS' OCCUPATION
25	Male	Married: 2 years Single at present	1 younger brother 1 younger sister 1 older sister	Mother: completed sixth grade Father: completed tenth grade	Father: Assembler

TEACHER PREPARATION	NUMBER DIFFERENT SCHOOLS ATTENDED		DEGREE OF SELF-SUPPORT		JOBS HELD FOR PAY	
Major: English	Elementary:	1	High school:	20%	Professional	1
Minor: French	Secondary:	1	College:	75%	Saleswork	3
Mathematics	College:	1	Since college:	100%	Factory	3
					Manual	3

NUMBER TRIPS TAKEN (NON-MILITARY)		VETERAN		EXPERIENCE OF YOUTH LEADERSHIP
Over 500 miles:	4	Domestic Service:	21 months	
Foreign country:	1	Foreign service:	6 months	
		Branch: Army		
		Rank: Staff Sergeant		

SUBJECT NUMBER 26

AGE	SEX	MARITAL STATUS	FAMILY DATA	PARENTS' EDUCATION	PARENTS' OCCUPATION
25	Male	Single	3 older brothers 1 younger sister 1 older sister	Mother: completed tenth grade Father: completed eleventh grade	Mother: seamstress Father: deceased 1941

TEACHER PREPARATION	NUMBER DIFFERENT SCHOOLS ATTENDED		DEGREE OF SELF-SUPPORT		JOBS HELD FOR PAY	
Major: History	Elementary:	1	High school:	15%	Professional	1
Economics	Secondary:	1	College	15%	Sales work	2
Minor: English	College:	1	Since college:	100%		

NUMBER TRIPS TAKEN (NON-MILITARY)		VETERAN		EXPERIENCE OF YOUTH LEADERSHIP	
Over 500 miles:	3	Domestic service:	6 months	Class offices:	2
		Foreign service:	18 months	Activity manager:	2 months
		Branch: Army			
		Rank: Staff Sergeant			

dissatisfaction with the toilet facilities available, furniture in the building, and the faculty room with reference to the number, degree of privacy afforded and regulations for its use. The subject responded to only three items with a degree of satisfaction in regard to the school plant and they were the adequacy, competency, and cooperativeness of the custodial service.

This subject received a composite (X_O) teacher efficiency rating of 2.00, which places him 12th in the whole group of forty-five.

SUBJECT NUMBER 17

The analysis of the subject's responses to the various items of the Inventory (Table X) shows that dissatisfaction was expressed with sufficient frequency and with such degrees of intensity as to indicate a generally negative response. The mean total percent of dissatisfaction was -3.0. This teacher was more dissatisfied than satisfied with the following five categories in rank order: the school plant, -47.0 percent; administration, -28.0 percent; instructional materials, -27.0 percent; students, -6.5 percent; and the people of the community, -5.0 percent.

The items in regard to the school plant (Appendix B-1) to which the subject responded with a degree of dissatisfaction were: the freedom from hazards and the appearance of the school grounds, the school building in regard to type of construction, degree of modernness, adequacy of exits, and provision for ventilation. Further, dissatisfaction was expressed with rooms, toilet facilities, floors, and furniture of the building. In addition, dissatisfaction was expressed in the lack of provision for a faculty room, and the adequacy, competency and cooperativeness of the custodial service.

Degrees of satisfaction in regard to the school plant were expressed for the location of the school plant, adequacy of heating and lighting. The items in regard to the administration (Appendix B-1) to which the subject responded in a dissatisfactory degree were the general educational level of the School Board, the practices of the Superintendent, and the administrative policy of the Superintendent in regard to standards for promotion, standards for student eligibility, teacher research and experimentation in the use of democratic procedures. The items in regard to the administration to which the subject responded in a satisfactory degree were the educational philosophy of the Board of Education, and the rules and regulations promulgated by them.

The subject received a composite efficiency rating (X_O) of 1.75, which places him 6th in the whole group of forty-five.

SUBJECT NUMBER 22

The analysis of the subject's responses to the categories of the Inventory (Table X) shows that dissatisfaction was expressed with a sufficient number of items and with such degrees of intensity and frequency that the response would appear to indicate an overall negative response. The balance of responses to all the categories was such that this subject received a mean total percent of satisfaction which was 0.5. The highest dissatisfaction of this teacher was shown in the items of the school plant, -12.5 percent. The instructional materials and the school organization received the same degree of dissatisfaction respectively, -11.5 percent. The next two lowest categories were the particular position, and the faculty receiving scores of -4.0 and 0.0 percent respectively.

The items in regard to the school plant (Appendix B-1) to which the subject responded with a degree of dissatisfaction were the degree of modernness of the building, its provisions for ventilation, and adequacy of performance of heating and lighting. In addition, negative responses were given to adequacy of size, appropriateness of decoration, convenience and arrangement, and the freedom from disturbances of rooms in the building. The subject also responded with dissatisfaction with the condition of the floors, and the lack of a faculty room for privacy. This subject responded to a satisfactory degree to only one item in regard to the school plant which was cooperativeness of the custodial service.

The items to which the subject responded to as being dissatisfactory with reference to the instructional materials (Appendix B-1) in regard to the library were: its size, number of references, its adequacy for all areas, recency of books and materials and their variety and range of difficulty, and the suitability of periodicals. In addition, a negative response was registered for the poor and worn condition of textbooks. There were only two items, in reference to the instructional materials, for which this teacher indicated a degree of satisfaction and they were the up-to-dateness of text materials and the provision for class transportation for field trips.

The items with which this teacher was dissatisfied in reference to the school organization (Appendix B-1) were the vacation allowed, length of the class periods, the rigid use of the school grading system, and with the importance it holds for the pupils. In addition, this teacher registered dissatisfaction with the diversity and variety of extra curricular offerings, type of student control used in the whole school along with the frequency, purpose, responsibility for organiz-

SUBJECT NUMBER 22

AGE	SEX	MARITAL STATUS	FAMILY DATA	PARENTS' EDUCATION	PARENTS' OCCUPATION
22	Male	Single		Mother: completed eighth grade Father: completed twelfth grade	Father: Purchasing agent

TEACHER PREPARATION	NUMBER DIFFERENT SCHOOLS ATTENDED	DEGREE OF SELF-SUPPORT	JOBS HELD FOR PAY
Major: History Economics Minor: English	Elementary: 2 Secondary: 1 College: 1	College: 40% Since College: 100%	Professional 1 Manual 2

NUMBER TRIPS TAKEN (NON-MILITARY)	VETERAN	EXPERIENCE OF YOUTH LEADERSHIP
None		None

SUBJECT NUMBER 17

AGE	SEX	MARITAL STATUS	FAMILY DATA	PARENTS' EDUCATION	PARENTS' OCCUPATION
22	Female	Single	1 younger brother	Mother: completed college Father: completed college	Father: Buyer

TEACHER PREPARATION	NUMBER DIFFERENT SCHOOLS ATTENDED	DEGREE OF SELF-SUPPORT	JOBS HELD FOR PAY
Major: French Minor: English	Elementary: 1 Secondary: 1 Colleges: 2	College: 50% Since College 100%	Professional 1 Clerical 2 Saleswork 1

NUMBER TRIPS TAKEN (NON-MILITARY)	VETERAN	EXPERIENCE OF YOUTH LEADERSHIP
Over 500 miles: 3 Foreign country: 1		Tutoring others: 3 months Sunday school teaching: 9 months Class offices held: 2-1/2 yrs.

ing, and the value of school assemblies to the students. Also, dissatisfaction was indicated with reference to the type and amount of guidance available to students. Furthermore, this teacher was dissatisfied with the aids available to teachers for the guidance of students. The items to which this subject responded to a satisfactory degree in regard to the school organization were the time of the beginning and closing of the school year and the hour of opening of the daily sessions. Finally, this teacher expressed satisfaction with the provisions made for nutritional needs of the students in school.

This teacher received a composite teacher efficiency rating (X_0) of 3.25, which places him 36th in the whole group of forty-five.

In looking back over the ten case studies, it appears that two of the highly dissatisfied teachers also received high efficiency ratings. These two cases will be examined in somewhat greater detail. Looking at the particular items with which they expressed dissatisfaction, Teacher 26 gave the larger overall negative response of these two teachers, namely, -9.0. While teacher 17 had a negative score of only -3.0, she was highly dissatisfied with the school plant, administration, and instructional materials. In spite of his dissatisfaction with the administration, he secured the highest efficiency rating of all the ten teachers studied. Teacher 26 did not like the community, the school plant, the people in the community, nor the school organization.

If we examine the responses of the two teachers with the highest satisfaction scores and the highest efficiency scores, we find that Teacher 4 was dissatisfied with only one item, namely, the custodial service. Teacher 13 responded with dissatisfaction to no items. In general it would appear from these case studies and other data supplied in this investigation, that teachers are more satisfied than dissatisfied with the learning-teaching situation in which they find themselves, and generally they possess more of this quality than poor teachers. Teachers 17 and 26 as already indicated are exceptions to this statement.

Ordinarily it would be presumed, too, that good teachers would have some definite dissatisfactions. That is, one would expect good teachers to be definitely dissatisfied with what they consider important aspects of the situation, particularly those closely related to the learning-teaching situation. Not to be dissatisfied with such items would indicate, in many instances, incompetency.

A further study of the ten individuals reveals that the five most dissatisfied teachers taught in communities with populations of 200, 519, 956, 1190, and 2050, respectively, while the five most satisfied taught in communities with populations of 1444, 6849, 11088, 11867, and 42365. Four of the five most satisfied teachers worked in larger communities, and the five most dissatisfied teachers worked in smaller communities. This is what one would expect considering the lack of facilities in smaller communities. Two of the five small community teachers were given high efficiency ratings. They probably did find much with which to be dissatisfied.

The analysis of the five most satisfied and dissatisfied teachers would seem to indicate that the size of community may be related to satisfaction-dissatisfaction but not efficiency ratings.

In analyzing the pattern of responses of the four case study teachers who received superior efficiency ratings, it appears that they were more satisfied than the other case study teachers with other items of the Inventory.

SECTION IV

SUMMARY AND CONCLUSIONS

THE PURPOSE of this investigation was to ascertain the relationships of certain environmental factors to teaching success.

The subjects of the study were sixty seniors who had graduated in 1950 from a four-year course in the School of Education at the University of Wisconsin, and who had completed their first year of teaching.

The chief data gathering device employed in the investigation was a Teacher Situation Inventory consisting of two pages of directions and nine sections of items relative to the situation listed under the following categories: the community, the people of the community, the school plant, the school administration, the school organization, the instructional materials, the faculty, the students, and the particular position. The inventory also contained a general information sheet. The Inventory was mailed to sixty-five 1950 graduates; forty-nine filled out and returned the Inventory. Four returns were unusable because they were inadequately filled out. Fifteen of the 16 non-responding individuals were visited and interviewed. Information was also obtained from the superintendent or supervising principal under whom the teacher worked relative to the efficiency of each subject. The criteria of teaching efficiency used were an acceptability rating, the Wisconsin Adaptation of the M-Blank rating, the State Department of Education rating, and a peer rating. A composite (X_0) of the above four measures was also used, plus a composite used by Montross.

The data from the Inventory were tabulated and the mean satisfaction-dissatisfaction determined for all categories of the Inventory for each respondee. These data were studied in relation to the various criteria of teacher efficiency.

Analysis of selected items was also made in regard to the reasons for the subjects' liking or disliking various aspects of the environment in which they found themselves. The responses of the fifteen most satisfied and the fifteen most dissatisfied teachers were examined in some detail in relation to efficiency ratings. An analysis of the responses to selected items of the Inventory obtained from interviewing fifteen people who did not respond to the form was made. This was done to determine how those not returning the Inventory responded to aspects of their environment. Finally, case studies of the five most satisfied and the five most dissatisfied teachers were made to explore in more detail the relationships between the reactions of the teachers to the school-community environonment and their efficiency.

A number of environmental factors were found to be related to teacher efficiency. The findings are summarized relative to the four questions asked earlier:

1. With what situational factors do teachers appear to be more satisfied or dissatisfied?

Teachers appear to be satisfied with: the administration, faculty, particular position, school organization, the community, instructional materials, school plant, people of the community, and students, in the order given. Sixteen teachers were satisfied with all nine factors. Five teachers were dissatisfied with six or more factors. The remaining twenty-four generally showed dissatisfaction with one or two factors. The two situational factors with which dissatisfaction was indicated most frequently were the students and the people of the community. With respect to all nine factors, the mean responses of all teachers showed a degree of satisfaction.

2. What situational factors seem to be associated with teaching success?

Low positive correlations were obtained between each of the efficiency ratings and the mean responses of all teachers to various aspects of the environment. Positive correlations of 0.20 or higher were found between an efficiency rating and the overall responses to the Inventory and for eight of the nine categories included in the Inventory. The analysis of the fifteen most satisfied and the fifteen most dissatisfied teachers' responses showed that the most dissatisfied group received lower satisfactory efficiency ratings than the most satisfied group on the M-Blank. The categories of the environment which appear to be more closely related than others to efficiency

the students, the faculty, and the school organization. The data with reference to the other two criteria were inconsistent.

3. To what degree can teacher success be predicted on the basis of situational factors?

While there is a positive relationship between efficiency and teacher satisfaction, the correlations are uniformly low, and taken as a group, are not likely to prove of substantial predictive value. In combination with other factors in teacher efficiency one would expect to reduce the error of estimate by adding to the regression equation a situational factor.

4. What situational factors do teachers think should be changed?

Some teachers were satisfied with nearly all aspects of their situations; others were dissatisfied with nearly all aspects of their situation. The following observations seem warranted:

a. The total group of respondees were least satisfied with their students and would like a better attitude from students toward learning, improved scholarship, consistent attendance, better attitudes toward the faculty, and greater desire to go to college.
b. Almost as many respondees were dissatisfied with the people of the community as those satisfied. They preferred a higher general educational level for the people, better everyday language usage, and greater interest in the arts.
c. Many respondees desired a better school plant. Those aspects of the school plant which they would like improved included a room for faculty members, more privacy, and improved personnel and services.
c. The principal improvements suggested in the category related to instructional materials were a larger library and more adequate library facilities for all areas of instruction, more up-to-date materials and a greater variety of classroom aids, better provisions for conducting field trips and more extensive use of community resources.
e. The major improvements desired with reference to community included better cultural and leisure time opportunities, provision for adult education, opportunity for employment, and more adequate bus and train schedules.
f. Those aspects of school organizations which received the greatest attention with suggestions for improvement were more and better organized school assemblies, better counseling and guidance services with a skilled person to be employed in

more schools and more guidance materials made available to teachers, teacher meetings in which the problems of instructors are given adequate attention, and clearly defined operating procedures for the student council.

g. Some suggestions were made in relation to the particular position for increasing salaries, improving the system of promotion, decreasing the number and size of classes, and decreasing the number of extra-curricular activities assigned to a given teacher.

h. The number, type and competence of assistants to the faculty, the adequacy of the professional library, and the method, purpose, and basis of evaluation of the faculty received many suggestions for improvement.

i. Suggestions were made for improved standards of promotion, for clarifying and improving teacher tenure regulations, and for building better administrator-teacher relations in planning and policy making.

Inasmuch as environment constitutes such an essential part of the field of operations for the education of youth and adults alike, it would seem that persons concerned with teacher education at the preservice and at the inservice level might well direct more attention to the discovery of aspects of the situation in need of improvement. The school administrator might use such information in the management of teaching personnel with particular reference to the environmental conditions within the school community. The teacher education institutions might incorporate such information in organizing the required education courses to acquaint prospective teachers with environmental conditions existing in the field that might limit or enhance their effectiveness. Such information should also enable teacher education institutions to help teachers build understandings, skills, and attitudes necessary to meet various environmental conditions. Teacher education institutions might also utilize findings from this study at the graduate level in courses for supervisors and administrators. Finally, the information should be of value to placement officials in that the abilities of teaching candidates and demands of situations might be considered more intelligently.

BIBLIOGRAPHY

1. Anderson, Harold M. A Study of Certain Criteria of Teaching Effectiveness, Doctoral Dissertation, University of Wisconsin, 1952.
2. Asch, Solomon E. Social Psychology (New York: Prentice-Hall, Inc., 1952), chs. 18, 19.
3. Barr, A. S. "The Measurement and Prediction of Teaching Efficiency: A Summary of Investigations," Journal of Experimental Education, XVI (June 1948), pp. 203-283.
4. Barr, A. S. and others. Educational Research and Appraisal (Chicago: J. B. Lippincott Co., 1953).
5. Brookover, Wilbur B. "The Relation of Social Factors to Teaching Ability," Journal of Experimental Education, XIII (June 1945), pp. 191-205.
6. Cook, Lloyd Allen and Elaine Forsythe Cook. A Sociological Approach to Education (New York: McGraw-Hill Book Co., 1950), ch. 18.
7. Domas, Simion J. and David V. Tiedeman. "Teacher Competence: An Annotated Bibliography," Journal of Experimental Education, XIX (December 1950), pp. 101-218.
8. Erickson, Harley E. A Factoral Study of Teaching Ability, Doctoral Dissertation, University of Wisconsin, 1952.
9. Garnett, Francis Campbell. A Realistic Philosophy of Religion (New York: Harper and Brothers, 1942).
10. Garrison, Karl C. "Comparative Responses of Teachers and Student Teachers to Various Items on the Teaching Situation Test," Elementary School Journal, XLV (February 1945), pp. 334-339.
11. Hartman, George W. "Some Assumed Determiners of Teaching Attitudes," Journal of Educational Research, XXXI (January 1938), pp. 357-362.
12. Hartman, George W. Educational Psychology (Chicago: American Book Co., 1941), pp. 217-223.
13. Klausmeier, Herbert J. Principles and Practices of Secondary School Teaching (New York: Harper and Brothers, 1953).
14. Krech, David and S. Crutchfield. Theory and Problems of Social Psychology (New York: McGraw-Hill Book Co., 1948), ch. 5, 6, 7.
15. Krug, Edward A. Curriculum Planning (New York: Harper and Brothers, 1950), ch. 6.
16. Lewin, Kurt. Principles of Topological Psychology (New York: McGraw-Hill Book Co., 1930).
17. Lins, Leo J. "The Prediction of Teaching Efficiency," Journal of Experimental Education, XV (September 1946), pp. 2-60.
18. Martindale, Frank S. "Situational Factors in Teacher Placement and Success," Journal of Experimental Education, XX (December 1951), pp. 1-77.
19. McClusky, Howard Y., and Floyd J. Strayer. "Reactions of Teachers to the Teaching Situation: A Study of Job Satisfaction," School Review, XLVIII (October 1940), pp. 612-623.

20. Montross, Harold W. A Study of Certain Aspects of Temperament and Success in Teaching, Doctoral Dissertation, University of Wisconsin, 1952.
21. Murphy, Gardner. Personality: A Biosocial Approach to Origins and Structure (New York: Harper and Brothers, 1947).
22. Peck, Leigh A. "A Study of Adjustment Difficulties of a Group of Women Teachers," Journal of Educational Psychology, XXVII (September 1936), pp. 401-416.
23. Rothney, John W. M. Counseling the Individual Student (New York: William Sloane Associates, 1949).
24. Simon, Donald L. "Personal Reasons For the Dismissal of Teachers in Smaller Schools," Journal of Educational Research XXIX (April 1936), pp. 585-588.
25. Symonds, Percival M. "How Teachers Solve Personal Problems," Journal of Educational Research, XXXVIII (May 1945), pp. 641-652.
26. Thomas, W. I. "Social Behavior and Personality," in Theory and Social Research, Edited by Edmund H. Volkart (New York: Social Science Research Council, 1951).
27. Thomas, W. I. "The Behavior Pattern and the Situation," Publications of the American Sociological Society (Chicago: University of Chicago Press, 1928).

ANALYSIS OF CERTAIN ASPECTS OF TEACHING ABILITY

MARGARET LOIS JONES
Illinois State Normal University
Normal, Illinois

SECTION I

THE PROBLEM

Introduction

MANY STUDIES have been made of graduates of teacher training curricula to determine their success as teachers. These studies have been concerned in one way or another with the measurement and evaluation of teaching success. Such evaluation is of considerable interest because of the importance of the teacher in the environment of the child. Baxter (6) points out: "It is no longer enough that the teacher be the possesser of knowledge. Today's teacher must be a 'social engineer', capable of setting up a provocative environment for children's learning, charting the course of each individual child through the ever-changing social relationships in which he is involved and assisting each pupil to grow in his understanding of himself and others." Considering this role of the teacher, it can be readily realized that it is of greatest importance to secure the best possible teachers for our schools.

Teacher training institutions are confronted not only with the problem of selecting those candidates who show the most promise of becoming successful teachers, but also the problem of providing those curricular and extra-curricular activities which will aid those candidates when they have entered the training program. Evaluation of teaching efficiency and identification of the factors which contribute to teaching success are important in both the selection and training of candidates.

Evaluation of teaching success has been approached in many different ways. One approach has been to ask students to list traits or characteristics of the teacher "liked most" and the teacher "liked least". Another approach has been to measure certain characteristics of prospective teachers in training and to correlate these measures with various estimates of success obtained after they had been taught.. A third approach has been to study good and poor teachers at the time they are teaching for the purpose of noting qualities or procedures associated with success.

Statement of the Problem

This dissertation is concerned with measures of certain characteristics of good and average teachers obtained at the time they are already teaching. Some preservice achievement measures will be used, but the study is primarily concerned with temperament as measured by objective tests and other measures.

A group of good teachers and a group of average teachers will be defined in terms of three criterion estimates of teaching success: Practice teaching grade, Placement Bureau Rating, and Principal's M-Blank Rating.

The primary purpose of this study is to determine whether a group of good teachers can be differentiated from a group of average teachers in terms of certain measures of preservice achievement, temperament and personality. In order to determine whether a pattern of these characteristics may differentiate the two groups, a multivariate analysis of variance technique will be employed.

Depending on the results of the analysis of variance, a further investigation will be made to determine the relationship between certain of the measured characteristics of teachers and a composite of the criterion estimates. The canonical correlation method proposed by Hotelling (33) will be employed to determine the weights to apply to each of the criterion estimates to form the composite.

The study attempts to answer the following specific questions:

1. Is there a significant difference between the group of good teachers and the group of average teachers in terms of the profile of group means for all the variables?
2. Is there a significant difference between the group of good teachers and the group of av-

---

*The author wishes to express sincere appreciation to the many persons who helped in the completion of this study: To Dr. A. S. Barr, major professor, sincere gratitude is expressed for the constant encouragement, kindness and guidance given during the course of this investigation; to Dr. Burton W. Kreitlow for his suggestions in the initial stages of this study and for his critical reading of the manuscript; to Dr. Chester W. Harris who helped in planning the statistical design and who made valuable suggestions on the organization of the material in the dissertation; and finally, to the teachers and principals who served as subjects and contributed willingly to this research.

erage teachers in terms of classes of variables (preservice achievement variables, temperament variables and personality variables)?

3. What is the canonical correlation between a weighted composite of the criterion estimates and a linear combination of certain of the preservice achievement, temperament and personality measures?

It appeared from the results of many of the researches in the area that measures of preservice achievement, temperament and personality of teachers might be promising in a study of the evaluation of teaching success.

In this investigation, the preservice achievement measures are those measures obtained for each teacher some time prior to her receiving the Teacher's Certificate and accepting a teaching position. Two preservice achievement measures were obtained from the American Council on Education Psychological Examination administered to each teacher at the time she entered the University of Wisconsin. Two other preservice achievement measures were based on (1) grades earned in professional courses in Education taken during the Junior and Senior years, and (2) grades earned in courses taken in the teaching major.

Temperament is investigated in this dissertation by means of two kinds of measures: (1) alternation tests of flexibility, and (2) creative effort tests of disposition rigidity. Both these kinds of tests are what have been called by Cattell (10a) and others "objective tests of temperament." These objective tests demand no attempt at self-evaluation by the subject, but are speed tests designed to measure ability of the subject to perform some task in a given situation.

Flexibility and disposition rigidity are two components of an aspect of temperament which has been termed perseveration by many writers. Perseveration has many theoretical connotations, such as Spearman's "mental inertia" concept (57) and the "memory phenomenon" interpretation by other writers (25). The present study avoids these connotations by accepting Cattell's (17) definition of disposition rigidity and Kleemeier and Dudek's (38) definition of flexibility.

Cattell defines disposition rigidity as a "difficulty (slowness) in turning from old to new responses to a situation when the individual wills to make them." This temperament factor is best measured by creative effort tests in which the individual tries to perform an old task in a new way. For example, it could be measured by the number of times the subject could write the numbers "237" in the usual way divided by the number of times he could write the same numbers using reverse strokes.

Flexibility is defined by Kleemeier and Dudek as the ability to shift from one task to another. They say this is best measured by alternation tests in which the subject is required to do first one simple task, then another, and finally is required to do both tasks in the same test. The third test is considered the measure of flexibility because the subject is required to shift from one task to another in the same test. For example, the subject might be required in Test 1 to answer as many two-digit addition problems as he could in two minutes, then in Test 2 to answer as many two-digit subtraction problems as he could in two minutes, and finally in Test 3 to answer as many mixed addition and subtraction problems as he could in two minutes. A high score on Test 3 indicates a higher degree of flexibility in shifting from one task to another.

Personality is investigated in this dissertation by a self-evaluation made of five "traits" chosen from a personality inventory.

Certain considerations regarding the meaning of the terms personality and temperament as used in this study are important. Allport (1) speaks of temperament as a certain class of raw material from which personality is fashioned. "Strictly speaking there is no temperament apart from personality, nor any personality devoid of temperament."

The term temperament is usually employed in speaking of dispositions that are almost unchanged from infancy throughout life.

Allport (1) gives this definition of temperament: "Temperament refers to the characteristic phenomena of an individual's emotional nature, including his susceptibility to emotional stimulation, his customary strength and speed of response, the quality of his prevailing mood, and all peculiarities of fluctuation and intensity in mood; these phenomena being regarded as dependent upon constitutional make-up, and therefore largely hereditary in origin."

Significance of the Study

Teacher training institutions attempt in various ways to evaluate candidates for the purpose of selecting those persons who are most likely to become good teachers. Knowledge of factors in teaching success is important in order to increase the effectiveness of these evaluations.

As a result of the present study, differences between good and average teachers, as defined by three criterion ratings, may be indicated or not indicated; in addition, relationships between teaching success and measures of certain teacher characteristics may be indicated or not indicated.

If differences are indicated or if significant relationships are found, this would contribute to the understanding of factors associated with teaching success. Results of the study should also provide a valuable source for pointing out teacher characteristics which should be investigated in further research.

If differences are not indicated, or significant relationships are not found, it must be concluded that in subsequent research other definitions, measuring devices or general plan of investigation should be employed.

SECTION II

DESIGN OF THE STUDY

The Teacher Population Studied

THE TEACHER population studied consisted of the 1951, 1952 and 1953 women graduates of the University of Wisconsin who had received the Teacher's Certificate from the School of Education and who were still teaching in Wisconsin secondary schools.

The population was further defined to include all those women who were teaching in secondary schools located within approximately one hundred miles of the City of Madison. One hundred and twelve persons were in this group.

From this group of one hundred and twelve persons, the names of sixty persons were randomly drawn. The Fisher-Yates Table of Random Numbers (20) was used in this procedure.

Although this randomly drawn sample consisted of sixty persons, the group finally studied in this investigation consisted of only forty-six persons for whom complete data were available. By the fall of 1954 when the tests used in this study were given, seven members of the sample of sixty had changed their residence from the State of Wisconsin. An additional seven persons had to be eliminated from the sample because two subscores (Quantitative Score and Linguistic Score on the American Council on Education Psychological Examination) used as part of the data in this study, were not available for them. Thus fourteen persons were eliminated from the randomly drawn sample of sixty.

Since fourteen persons were eliminated from the sample of sixty, the group of forty-six women teachers finally studied cannot be considered precisely a randomly drawn sample. In order to estimate the amount of bias which may have resulted from this loss, certain comparisons between the group of forty-six teachers studied and the group of fourteen teachers not studied would seem pertinent. If the two groups are similar in terms of these comparisons, the effects of the bias may be judged a minimum. This would provide some evidence for the use of the sample of forty-six persons in a statistical design requiring the assumption of random sampling.

Comparisons between the group of forty-six and the group of fourteen were made in terms of (A) the Mean Total Grade Point Average for each group and (B) the Median Percentile Rank (total score) for each group on the American Council on Education Psychological Examination. Table I shows these comparisons.

A test of significance of difference between means, given by McNemar (43), was used to test the significance of difference between Mean Grade Point Averages. A sum of ranks test (61), applicable to non-normal distributions, was applied to test the hypothesis of no difference between the two Median Percentile Ranks.

Neither of these differences was statistically significant. Therefore, it would seem the group of forty-six and the group of fourteen were similar in terms of academic achievement as measured by grade point average and intelligence as measured by the American Council on Education Psychological Examination. It is believed the sample of forty-six, while not precisely a randomly drawn sample, may be used in an analysis of variance design.

Description of the Subjects

The forty-six subjects in this investigation were women teachers in secondary schools in Southern Wisconsin. They were all graduates of the University of Wisconsin and had received the Teacher's Certificate from the School of Education. Six members of the group had graduated in 1951, seventeen in 1952 and twenty-three in 1953. The range in ages to nearest birthday was from 22 to 26 years while the average age for the group was approximately 24 years. Nineteen of the group were married and twenty-seven were not. Thirty-four entered the University of Wisconsin as freshmen and completed the entire four year college course at this University. Twelve entered the University as transfer students.

Table II shows the variety of subjects taught by the forty-six teachers.

Certain comparisons between the forty-six women who are subjects in the present study and women graduates of the University of Wisconsin School of Education in previous years are shown in Table III.

The data in Table III would seem to indicate the group in the present study is not markedly different from women graduates of other years in terms of academic achievement.

The Mean Total Score Percentile Rank on the American Council on Education Psychological Examination for the group of forty-six teachers is 62.46. Lins (41a) obtained a mean of 61.42 for the fifty-eight (1942) women graduates who were subjects in his study.

TABLE I

COMPARISON BETWEEN THE GROUP STUDIED AND THE GROUP NOT STUDIED IN TERMS OF TOTAL GRADE POINT AVERAGE AND TOTAL SCORE MADE ON AMERICAN COUNCIL ON EDUCATION PSYCHOLOGICAL EXAMINATION

	Mean Total G.P. Average	Med. %ile Rank (Total Score) ACE
Group of 46 teachers studied	1.92	67
Group of 14 teachers not studied	1.89	63

TABLE II

SUBJECTS OR COMBINATIONS OF SUBJECTS TAUGHT BY THE FORTY-SIX TEACHERS

Subject or Subjects Taught	Number Teachers Teaching Each Subject or Combination
English, English Literature, or English plus another subject	15
Math and Science	4
Home Economics	16
Foreign Language (French, Latin, Spanish)	2
History, Civics, Social Studies	1
Art Education	3
Business Education, Commerce	1
Music	4

TABLE III

COMPARISON BETWEEN SUBJECTS IN PRESENT STUDY AND GRADUATES OF PREVIOUS YEARS

	Subjects in Present Study N = 46	1942 Women Graduates (Lins' Study) N = 58	1950 Women* Graduates N = 136
Total Grade Point Average			
Mean	1.92	1.877	1.889
S. D.	0.38	0.382	0.359

*Data obtained from the Office of Student Personnel Statistics, University of Wisconsin, Madison, Wisconsin.

TABLE IV

TEST-RETEST RELIABILITY ESTIMATES OF KLEEMEIER-DUDEK BATTERY OF FLEXIBILITY TESTS (N = 53)

Name of Test	Test-Retest Reliability Estimate
Two-Digit Numbers	
Addition	.78
Subtraction	.87
Mixed	.88
Word Completion	
Add Final Letter	.59
Add Initial Letter	.71
Mixed	.53
Perceptual Speed	
Count N's	.70
Count W's	.72
Mixed	.75

The group presently studied appear to be reasonably representative of women graduates of previous years who have received the Teacher's Certificate from the University of Wisconsin School of Education.

Data-Gathering Devices Employed

Twenty-three variables were involved in the present study. Three of the variables were estimates of teaching success and twenty were measures of certain teacher characteristics. Of these twenty measures of teacher characteristics, four were measures of preservice achievement, five were measures of personality and eleven were objective tests designed to measure certain aspects of temperament.

The twenty-three variables are listed below. The number preceding each variable is used to identify that measure in reporting the analysis of the data. A code symbol follows each measure.

Measures of Preservice Achievement

(1) Quantitative Score (%), American Council on Education Psychological Examination, 1948 edition	Q
(2) Linguistic Score (%), American Council on Education Psychological Examination	L
(3) Professional Grade Point Average, based on grades earned in Education 73, 74, and 75	PGPA
(4) Grade Point Average earned in courses taken in teaching major	MGPA

Objective Measures of Temperament

Creative Effort Tests—Cattell

(5) Reverse Strokes Test	Disposition Rigidity	Sub 6
(6) The Alphabets Test	Disposition Rigidity	Sub 7

Battery of Alternation Tests—Kleemeier-Dudek

(7) Two-Digit Numbers, Addition	TDN Add
(8) Two-Digit Numbers, Subtraction	TDN Sub
(9) Two-Digit Numbers, Mixed Flexibility	TDN Mix
(10) Word Completion, Add final letter	WC Final
(11) Word Completion, Add initial letter	WC Initial
(12) Word Completion, Mixed Flexibility	WC Mix
(13) Perceptual Speed, Count N's	PS N's
(14) Perceptual Speed, Count W's	PS W's
(15) Perceptual Speed, Mixed Flexibility	PS Mix

Guilford-Zimmerman Temperament Survey

(16) General Activity vs., Slowness Inactivity	G
(17) Restraint vs., Impulsiveness	R
(18) Ascendance vs., Submission	A
(19) Sociability vs., Shyness, Seclusiveness	S
(20) Emotional Stability vs., Neuroticism	E

Criterion Estimates of Teaching Success

(21) Practice Teaching Grade, Education 75	PT
(22) Placement Bureau Rating	PB
(23) Principal's M-Blank Rating	PM

Sources of Data

The following data were obtained from University records:

1. Quantitative Score and Linguistic Score were obtained from records kept at the Student Counseling Center of the University. These scores were derived from the American Council on Education Psychological Examination which is administered to all students entering the University.
2. The Grade point averages earned in the three professional Education courses and in the courses taken in the major teaching field were obtained from records in the Registrar's Office of the University.

Data for assessing temperament was obtained by administering the two creative effort tests and the Kleemeier-Dudek Battery of alternation tests to each of the forty-six subjects. Two measures of temperament were obtained from the creative effort tests and three measures of temperament were obtained from the battery of alternation tests.

Data for assessing personality was obtained by administering the Guilford-Zimmerman Temperament Survey to each teacher. Scores on only five of the ten traits assessed by this inventory were included in the data for this study.

An estimate of the success of each teacher was obtained from the principal of the school in which the teacher was currently teaching. A letter explaining the study and enclosing a copy of the Wisconsin Adaptation of the M-Blank (Appendix B)* was sent to each principal.

---

*All references to Appendices may be found in original thesis, Library, University of Wisconsin.

The practice teaching grade for each subject was obtained from records on file in the Wisconsin High School Office of Student Teacher Records.

The Placement Bureau rating was obtained from the Assistant Director of the University of Wisconsin Teacher Placement Bureau.

Description of the Measures Employed

(1) American Council on Education Psychological Examination for College Freshmen (Appendix C)

Percentile rank on each of two subscores Quantitative and Linguistic, yielded by this examination were measures of intelligence employed in this study. Fourteen of the subjects who entered the University in the Fall of 1948 were administered the 1947 edition of the test; the remainder of the subjects (32) who entered in the Fall of 1949 or as transfers in 1950, were administered the 1948 edition of the test. Percentile ranks were determined on the basis of national norms.

The examination consists of six tests grouped in two general classes as follows:

Quantitative Tests (Q Score):	Linguistic Tests (L Score):
Arithmetical Reasoning	Same-Opposite
Number Series	Completion
Figure Analogies	Verbal Analogies

Subscores based on results of these two sets of tests were used in the present investigation. Factor analysis studies have indicated that two indistinct kinds of mental abilities appear to be involved in performance on these six tests (19). High Q and L scores indicate high levels of performance in each of the abilities assessed.

(2) Professional grade point average

This average was computed on the basis of grades earned in Education 73, 74 and 75. These are professional courses in education offered by the School of Education at the University of Wisconsin. A high grade point average indicated a higher level of academic achievemnt in professional education courses.

(3) Grade point average in the major teaching field

All candidates for the Teacher's Certificate are required to take a certain number of credits in a major teaching field. An average of the grades earned in these courses was computed for each subject. Professional education courses were excluded from this average, as well as the first two years of basic courses required of all freshmen and sophomores. A high grade point average indicated higher academic achievement in the major teaching field.

(4) Creative Effort Tests of Disposition Rigidity (Ref., 13. Also Appendix C 4)

1. Reverse Strokes—This test consisted of two parts: In Part A the subject wrote the numbers "237" as many times as he could in one minute. In Part B, the subject wrote the same numbers for one minute, using a reversed stroke, i. e., beginning at the bottom of the "7" and ending at the top of the "2". The score was the number of digits written correctly in Part A divided by the number of digits written correctly in Part B. Since a low score on this test indicated a higher degree of ability to overcome rigidity, these scores were reflected in order to be consistent with scores on the remaining variables.

2. Alphabet Test—This test consisted of three parts. In each part, the subject was to write as many letters as he could in the time allotted, as follows:

Part A, the letters "abcd" were to be written	30 seconds
Part B, the letters "ABCD" were to be written	30 seconds
Part C, the letters "aAbBcCdD" were to be written	30 seconds

The final score was the number of letters written correctly in Part A plus Part B divided by twice the number written correctly in Part C. As in the Reverse Strokes Test, a low score indicated better performance; therefore the scores made on this test were reflected in order to be consistent with the remainder of the data.

(5) Battery of Alternation Tests of Flexibility (Ref., 38. Also Appendix C 3)

This battery consisted of a total of nine tests. There were three tests in each of three areas: Numerical Speed, Verbal Abilities, and Perceptual Speed. In each area, the subject was required to do first one task, then another task, and finally in the third test was required to shift from one task to another in the same test. This third test was considered the measure of flexibility. Thus three measures of flexibility were obtained from the battery of nine tests:

1. Flexibility in Numerical Abilities
2. Flexibility in Verbal Abilities
3. Flexibility in Perceptual Speed

All tests in the battery were speed tests and were administered in two parts so that estimates of test-retest reliabilities could be computed. Answers to the tests were recorded on IBM sheets. Since many of the subjects were tested individually, a set of standard directions (Appendix D) was used each time the tests were administered.

The tests in each of the three areas are described as follows:

1. Two-Digit Numbers Test (Numerical Abilities)

Each of the three tests in this area consisted of 120 items administered in two parts of 60 items each. The time limit for each part was 90 seconds. The subject's task was to indicate on an IBM answer sheet whether answers to the problems as given were right or wrong. In the first test the subject responded to addition problems and in the second test the subject responded to subtraction problems. The third test consisted of mixed items, i. e., both addition and subtraction problems. In responding to the mixed items on this test the subject was required to shift from one task to another.

A high score on the third test is an indication of a higher level of flexibility in the area of numerical abilities.

2. Word Completion Tests

This series consisted of 3 tests of 60 items each. Each test was administered in two parts of 30 items each with a time limit of 90 seconds for each part. The subject's task was to select the one letter from among five alternatives which formed a word when used with a given stem of three letters. None of the stems were three letter words in and of themselves.

In the first test in this series, the subject was to find the letter which, added to the front of the stem, made a complete word. In the second test, subject was required to find the letter which, added to the back of the stem, made a complete word. In the third test, the tasks were mixed. The letters might go either in front or in back of the stem. No cues other than the stem itself were given as to whether the answer was an initial or final letter.

A high score on test 3 was an indicator or a higher level of flexibility in the area of verbal abilities.

3. Perceptual Speed Tests

Each of these tests consisted of 60 items administered in two parts of 30 items each. The time limit for each part was two minutes. Each item consisted of a line of 30 capital block letters. The subject's task was to count the number of times some particular letter appeared. This letter appeared from one to five times in each line. The number of times it occurred was entered directly on the IBM answer sheet.

In the first test in this series, the subject counted the number of N's appearing in each line. In the second test, subject counted the number of W's appearing in each line. In the third test, each line consisted of ''M's'' ''N's'' and ''W's''. At the beginning of each line, the letter to be counted was indicated in parenthesis. The requirements of which letter to look for on each line were mixed.

A high score on the third test was an indication of a higher level of flexibility in the area of perceptual speed.

Reliabilities of each of the tests in the battery of alternation tests is shown in Table IV.

These reliability coefficients were computed on the basis of scores earned on two separately timed administrations of each of the nine tests. Thorndike (42) discusses methods for computing reliabilities of highly speeded tests, as distinct from ''power'' or level of performance tests with unlimited time. He points out that reliability estimates for such speed tests should not be based on any division of a single administration of the test. The two ''part'' scores of the single administration would be very nearly identical and the resulting correlation between them would be misleadingly high.

(6) Guilford-Zimmerman Temperament Survey (Appendix C 2)

This questionnaire was developed to assess ten traits of personality. Scores on only five of these traits were used as measures of personality in the present study. These five were: General Activity vs., Inactivity, Slowness; Restraint vs., Impulsiveness; Ascendance vs., Submissiveness; Sociability vs., Shyness; Emotional Stability vs., Neuroticism.

Each trait is appraised in terms of thirty items. Items are stated affirmatively rather than in question form. The subjects responded to each item by marking on an IBM answer sheet under one of the following headings: ''YES'', ''?'', ''NO''. In the present study, the subjects responded to all items on this inventory using either the ''YES'' or ''NO'' answer.

Odd-even and split-half reliability estimates for each of the five personality traits were computed for 53 of the 60 persons in the original sample. These reliabilities are shown in Table V. The figures in parenthesis indicate the reliability of the entire test found by using the Spearman-Brown formula (29).

Table VI shows the intercorrelations of the scores on the five personality traits. In general

TABLE V

RELIABILITY ESTIMATES FOR FIVE TRAITS OF PERSONALITY MEASURED BY THE GUILFORD-ZIMMERMAN TEMPERAMENT SURVEY

Personality Trait	Odd-Even Reliability Estimate		Split-Half Reliability Estimate	
G	.70	(.82)	.55	(.71)
R	.64	(.78)	.48	(.65)
A	.52	(.68)	.58	(.73)
S	.72	(.84)	.66	(.80)
E	.68	(.81)	.69	(.82)

TABLE VI

INTERCORRELATIONS OF SCORES ON FIVE TRAITS OF PERSONALITY MEASURED BY GUILFORD-ZIMMERMAN TEMPERAMENT SURVEY

(N = 53)

	G	R	A	S	E
G					
R	.009				
A	.423	.102			
S	.430	.038	.435		
E	.012	.241	.172	.429	

the intercorrelations are low, indicating there is considerable unique contribution made by each trait. If all the intercorrelations had been high, the five separate scores could not have been used in this study on the grounds that only one personality trait was being measured. Since some of the intercorrelations were low, the scores were used as representing five traits of personality.

The five traits assessed by this instrument are described below. These descriptions are quoted from the Manual of Instructions and Interpretations for the Guilford-Zimmerman Temperament Survey (30).

1. General Activity vs. Inactivity, Slowness

A person scoring high on this trait usually likes a rapid pace of activities rather than a slow and deliberate pace. He prefers keeping in motion rather than pausing to rest; he has a liking for speed, hurrying and quickness of action.

2. Restraint vs. Impulsiveness

A person scoring high on this trait is likely to be restrained and serious in contrast to impulsive. Serious-mindedness, deliberateness, persistent effort and self-control are emphasized by this trait. A person scoring low is characterized as happy-go-lucky, carefree, impulsive, and excitement-loving.

3. Ascendance vs. Submissiveness

Persons with high scores in this area have strong leadership habits, enjoy speaking with individuals and speaking in public; they enjoy persuading others and being conspicuous. A high score indicates social boldness in contrast to submissiveness.

4. Sociability vs. Shyness, Seclusiveness

Persons with high sociability scores have many friends and acquaintances, readily enter into conversations, like social activities, and seek out social contacts. The connotation for people scoring low in this category is "shyness" and "seclusiveness".

5. Emotional Stability vs. Neuroticism

Persons with high score in this area usually are cheerful and optimistic and have evenness of moods, interests and energy. They are not excitable over slight distractions or annoyances, nor do they have perseveration of ideas and moods.

(7) Measures of the Criteria of Teaching Success

Before describing the measures employed as criteria, a brief discussion of certain terms will be presented.

The term "successful" as applied to the teacher has been used in the present study rather than the term "effective". The two terms often are, but need not be used synonymously.

The term "effective" has been used to describe a teacher who provides an atmosphere which fosters desirable growth in pupils. Using this term implies (1) that agreement would have to be reached on what changes in pupils were desirable, and (2) methods for assessing them would have to be determined.

Up to the present time, there is no adequate system for evaluating a teacher's "effectiveness" as defined above.

Success as used in this study, is a broader term and involves the teacher as a friend and counselor of pupils and a member of a community and professional staff. To be successful in this broader sense a teacher must not only be able to provide an atmosphere which fosters desirable growth in pupils but she must be acceptable to her peers, her pupils, the community and the principal of the school in which she teaches.

1. Practice Teaching Grade

These grades are based on two semesters of practice teaching experience, one semester in the major field and one semester in the minor field. Grades are usually awarded on a Superior-Average-Inferior basis using an A, B, C grading system. For purposes of statistical treatment, these grades were transformed to a value scale of A = 13, A- = 12, B+ = 11, B = 10, B- = 9, C+ = 8, C = 7, C- = 6.

2. Placement Bureau Rating

This rating was made by the Assistant Director of the Placement Bureau of the University of Wisconsin. The ratings were based on a variety of information collected by the Placement Bureau office: departmental rating of candidate, transcript, letters of recommendation, biographical statement by candidate, scores on personality and intelligence tests, and information obtained from personal interviews with each teacher. This information is compiled on one data sheet for each candidate.

The Placement Bureau official rated each teacher on a thirteen point scale: 1, 1-, 2+, 2, 2-, 3+, 3, 3-, ... to 5. These ratings were transformed to a value scale of 1 = 13, 1- = 12, 2+ = 11, 2 = 10, 2- = 9, 3+ = 8, ... 5 = 1.

3. Principal's M-Blank Rating Form (Appendix B)

On this form the principal of each teacher was asked to list strengths and weaknesses in the teacher's work and to assign an over-all rating on a five-point scale. The categories on a five-point scale were: 1 = superior; 2 = above average; 3 = average; 4 = below average; 5 = inferior. These ratings were transformed to the same value scale as that used for the other two criterion ratings: 1 = 13, 2 = 10, 3 = 7, 4 = 4, 5 = 1.

Procedures Followed in Treatment of the Data

Three criterion estimates and twenty measures of certain teacher characteristics are the basic data for this study. The three criterion estimates are: rating by principal, rating by Placement Bureau office, and practice teaching grade. The set of twenty measures of teacher characteristics consists of the following:

1. Four are measures of preservice achievement.
2. Five are measures of personality traits.
3. Two are measures of the disposition rigidity temperament factor obtained by administering the creative effort tests.
4. Three are measures of the flexibility temperament factor obtained by administering the battery of alternation tests.
5. Six of the measures were obtained from tests which were part of the battery of alternation tests.

Two methods are used in analyzing these data:

1. Differences between a group of good teachers and a group of average teachers, as defined by the three criterion estimates, are investigated by means of a multivariate analysis of variance technique (8). By this technique it is possible to determine whether the two groups differ in any or all of three ways:
   a. In terms of a sum of scores made by individuals on the twenty measures;
   b. In terms of classes of measures (preservice achievement measures, temperament measures, and personality measures).

2. Relationships between seven of the twenty measures of teacher characteristics and a composite of the three criterion estimates are investigated by means of the canonical correlation method given by Hotelling (32, 33). Seven variables, selected on the basis of profile differences found to exist between the group of good teachers and the group of average teachers, are included in this study of relationships.

The canonical correlation method was selected because it gives the solution for determining weights to apply to each of the three criterion ratings to combine them into a composite criterion. The weighted composite obtained by this method is a general factor extracted from the matrix of intercorrelations of the three criterion variables. As such, it is possible to describe the factor, or composite, in terms of the relative contribution of the variables.

Thus the advantage of a weighted composite is in being able to identify the variables associated with it. Holzinger (31), and Cumbee and Harris (19), have shown that a simple averaging of a set of criterion ratings may not result in all the ratings being represented in the composite. That a simple averaging of criterion estimates may not be feasible is also pointed out by the studies of Anderson (2) and Erickson (23).

In brief, the procedures to be followed in treating the data are:

1. Two groups of teachers will be formed by dividing the sample of forty-six into an upper half and a lower half on the basis of all three criterion estimates received by each teacher.
2. Raw scores on the twenty measures will be standardized over both groups.
3. A multivariate analysis of variance technique will be applied as a test of significance of differences between the two groups of teachers.
4. Certain variables will be selected from the twenty measures of teacher characteristics on the basis of the results of the analysis of variance. If the groups differ in terms of classes of variables, then classes of variables will be selected for the further study of relationships. If the groups differ in terms of profile, then variables which separate the groups the most will be selected.
5. The canonical correlation between a composite of the three criterion estimates and the selected measures of teacher characteristics will be computed. The weights to form the composite will be given by this computation.

SECTION III

ANALYSIS OF THE DATA—ANALYSIS OF VARIANCE TREATMENT

IN THIS CHAPTER the analysis of the data as treated by the multivariate analysis of variance technique will be presented.

The primary purpose of the present study was to investigate differences between a group of good teachers and a group of average teachers in terms of certain measures of teacher characteristics. While some pre-service achievement measures were employed, the teacher characteristics of most concern in the study were five personality traits and certain aspects of temperament as measured by objective tests.

The analysis of possible group differences was made by a method proposed by Block, Levine and McNemar (8). This was a multivariate analysis of variance design which permitted differences between the group of good teachers and the group of average teachers to be revealed in any or all of three ways:

1. In terms of sums of scores for individuals;
2. In terms of classes of variables (preservice achievement variables, personality variables, and temperament variables);
3. In terms of a profile of group means on twenty variables.

Forty-six women teachers were the subjects in the investigation. This sample was divided into two equal groups of twenty-three members each on the basis of the three criterion ratings. Those teachers who received the highest ratings were assigned to the upper half; those who received the lowest ratings were assigned to the lower half. All three ratings received by each subject were taken into consideration in separating the sample into the two groups.

Table VII shows the combinations of the three ratings assigned each of the forty-six teachers, and the number of teachers receiving each combination.

From Table VII it is noted that all teachers in the upper group received ratings of superior or above average from all three raters. The range in ratings received by any one person in this group is from 1 to 2. There appears to be agreement among the various raters that these persons are superior and above average teachers. In the present study, this group is designated the group of "good" teachers and is referred to in subsequent tables by the code symbol "Group I."

Table VII shows there is less uniformity among the ratings assigned the teachers in the lower group. The ranges in ratings received by any one person in this group were from 1 to 3, 2 to 3, 2 to 5, 3 to 4, and 3 to 5. One person in the lower group received one rating of superior and two ratings of average; ten persons received one rating of above average with two ratings of average, below average or inferior. Twelve members of the lower group are clearly average or below average teachers in terms of all three ratings. The lower half of the sample is designated the group of "average" teachers in the analysis which follows. The code symbol "Group II" is used to refer to this lower group.

The raw scores on each variable were summed over both groups. The mean and standard deviation of the distribution for each variable were computed. These means and standard deviations were then used in converting the raw scores to standard scores. Raw scores and standard scores for all variables are shown in Appendices E and F.

Standardizing the scores resulted in each of the twenty distributions having a mean of zero and a standard deviation of 1.0. Thus the variance for each variable was equalized.

Equalizing the variance in this way eliminated the variables as a source of variance. When the raw scores were converted to standard scores, the sum of scores on each variable became zero. Since these sums are the basis for computing the mean square deviation for variables, it is obvious that variables as a source of variance would be eliminated.

The standard scores for the two disposition rigidity measures, the Reverse Strokes test and the Alphabets test, were reflected. As indicated earlier in this study, a low score on these two tests indicated better performance while a high score on the remaining eighteen variables indicated the better performance.

The variance table for the analysis of the sums of scores made by individuals, Group I and Group II, and the analysis of the group means on the several variables, Group I and Group II, is presented in Table VIII. The significance of all F ratios found in this study was tested by tables in Edwards (22).

The appropriate test for determining whether or not groups differ in terms of the sum of scores on the twenty measures is the mean square for groups over the mean square for individuals.

This ratio is 11.17. For 1 and 44 degrees of freedom, an F of 7.25 is significant at the 1% level. The hypothesis that the two groups do not differ in their grand means, i.e., the means given by pooling the variables within each group, is therefore rejected. The group of good teachers and the group of average teachers are significantly different in terms of the sum of scores made by individuals on all twenty of the preservice, temperament and personality measures.

The separate group means for the several variables are the basis for possible group profiles. The appropriate test for determining whether the profiles differ significantly is the mean square for the group by variable interaction over the mean square for individuals by variable interaction. This ratio is 2.05. For 19 and 836 degrees of freedom an F of 1.89 is significant at the 1% level. The hypothesis that the groups do not differ in the profile of scores is therefore rejected. The group of good teachers is significantly different from the group of average teachers in terms of the profile of scores.

The profile of scores for each group is shown in Figure 1, next page. These profiles were plotted on the basis of the means of the distribution of standard scores for each group on each variable. Since the sum of the standard scores for Group I on each variable and the sum of the standard scores for Group II on each variable were very nearly identical, the absolute values

TABLE VII

COMBINATIONS OF CRITERION RATINGS ASSIGNED THE FORTY-SIX TEACHERS BY THREE RATERS

Combinations of the 3 ratings assigned to each teacher	Number of teachers who received each combination of ratings
1, 1, 1	5
1, 1, 2	1
1, 2, 2	12
2, 2, 2	5
1, 3, 3	1
2, 3, 3	9
2, 4, 5	1
3, 3, 3	10
3, 4, 3	1
3, 5, 3	1
Total Number	46

TABLE VIII

VARIANCE TABLE FOR ANALYSIS OF GRAND MEANS, GROUP I AND GROUP II, AND ANALYSIS OF GROUP MEANS ON THE SEVERAL VARIABLES, GROUP I AND GROUP II

Source	Sum of Squares	d.f.	Mean Square	F
Variables	0	19	0	
Groups	45.3146	1	45.3146	
Individuals	178.4694	44	4.0561	11.1720*
Group × Variables	30.9332	19	1.6281	
Individuals × Variables	664.9632	836	0.7954	2.0469*
Total	919.6160	919		

*Significant beyond the 1% level.

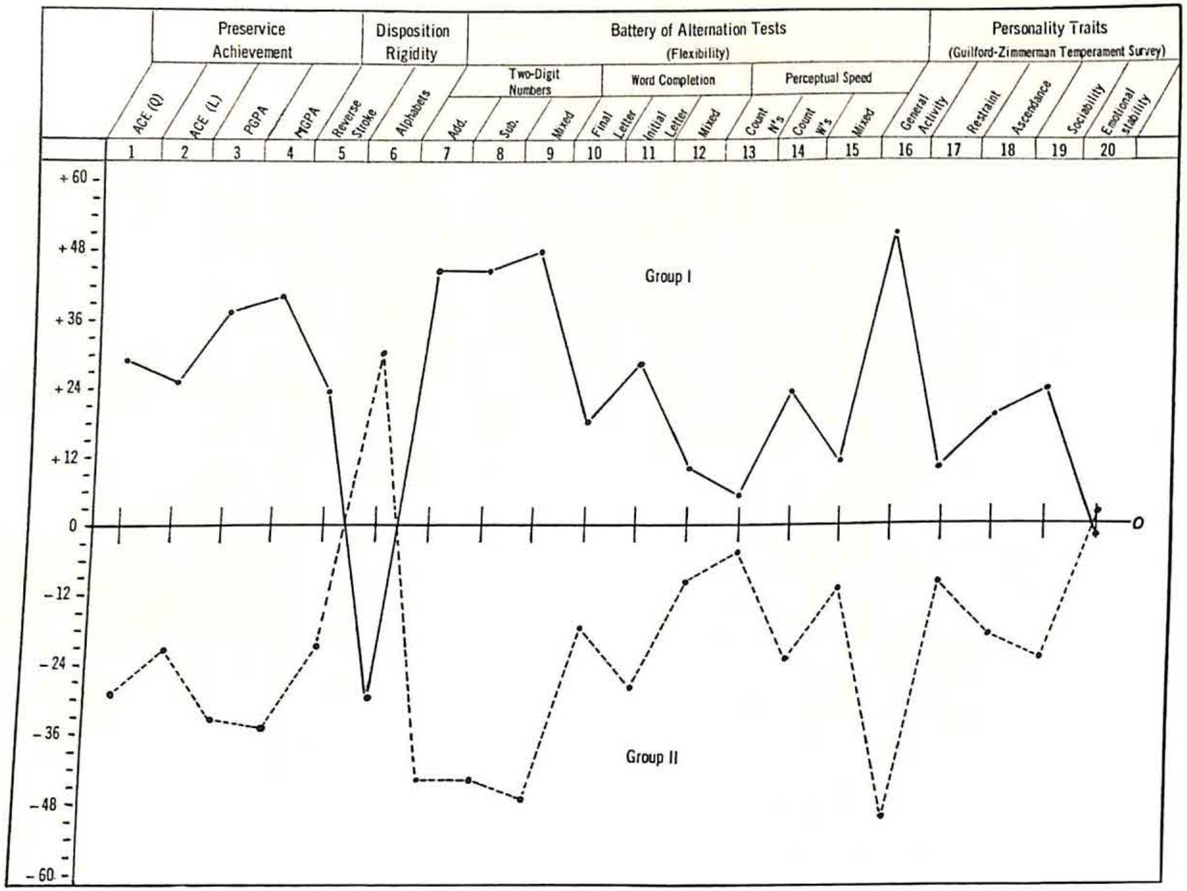


Figure 1. Profile of Group Means for Twenty Variables

of the means were the same for each group on each variable. Thus the profiles were different for the two groups only in terms of their sign, positive or negative. The important consideration in interpreting the profile differences was the distance the groups were separated on each of the variables.

Since a significant difference was found between the groups in terms of profiles, comparisons are made between the group of good teachers and the group of average teachers in terms of the twenty measures. These comparisons are:

1. There appears to be very little difference between the good teachers as compared to the average teachers in terms of emotional stability as measured by the Guilford-Zimmerman Temperament Survey.

2. There does appear to be a great deal of difference between good teachers and average teachers in terms of the General Activity personality trait.

3. Good teachers in comparison to average teachers tend to differ in certain other respects: professional grade point average, major teaching field grade point average, flexibility in numerical abilities and the disposition rigidity temperament factor as measured by the Alphabets Test.

Having found the profile accounted for significant differences between the two groups, the data were further analyzed to determine whether classes of variables (preservice achievement variables, personality variables, and temperament variables) would account for significant differences between the groups.

The solution to this problem involved separating the groups by variable source of variation and the individuals by variable source of variation each into two parts: between classes variation and within classes variation. Statistically this amounted to an extension of the first analysis of variance.

Interaction terms were computed with the variables now considered classes of variables.

Variable I	was comprised of the 4 academic measures.
Variable II	was comprised of the 11 objective measures of temperament.
Variable III	was comprised of the 5 personality measures.

The sums now required for computing the interaction terms were:

1. For groups by between classes—
   Group sums on all variables in each of the three classes (resulting in a 3 × 2 matrix)
2. For individuals by between classes—
   Sums of scores made by individuals on all variables in each of the three classes (resulting in a 3 × 23 matrix)

An individuals by between classes interaction term was computed separately for Group I and Group II, and the variances tested for homogeneity before pooling.

The group by between classes interaction was subtracted from the total group by variable interaction (shown in Table VIII) to give the groups by within classes interaction. Individuals by within classes interaction was obtained by a similar method.

A summary of the basic analysis of variance and the extension which analyzes classes of variables as a source of variation, is presented in Table IX.

The appropriate test for determining whether individuals vary significantly with classes of variables is the mean square for individuals by between classes interaction divided by the mean square for individuals by within classes. This ratio is 2.72. For 88 and 748 degrees of freedom an F of 1.49 is significant at the 1% level. The hypothesis of no difference is rejected. Therefore, individuals perform better on some classes of variables than on others.

The ratio of the mean square for groups by between classes interaction over the mean square for individuals by within classes is the appropriate test for determining whether groups differ in terms of classes of variables. This ratio is 1.09. For 2 and 748 degrees of freedom an F ratio of 4.62 is significant at the 1% level. The hypothesis that the groups do not differ in terms of classes of variables is accepted. Therefore, the interaction of groups and classes of variables is not significant.

The results of these last two tests of significance indicate that while there is some evidence individuals vary in terms of classes of variables, there is no evidence the two groups differ in terms of classes of variables.

Summary of Analysis of Variance Treatment of Data

Significant differences were found to exist between the group of good teachers and the group of average teachers in two ways:

1. In terms of the sum of scores made by individuals on the twenty measures;
2. In terms of the profile of scores.

There was some evidence individuals varied in terms of classes of variables but there was no evidence the groups differed in terms of clas-

TABLE IX

SUMMARY OF ANALYSIS OF VARIANCE

Source	Sum of Squares	d.f.	Mean Square
Variables	0	19	0
Groups	45.3146	1	45.3146
Individuals	178.4694	44	4.0561
Classes			
(Groups × Variables)	(30.9332)	(19)	(1.6281)
Groups × between classes	1.4701	2	0.7350
Groups × within classes	29.4631	17	1.7331
(Individuals × Variables)	(664.9623)	(836)	(0.7954)
Individuals × between classes	161.5263	88	1.8329
Individuals × within classes	503.6687	748	0.6734

TABLE X

EXTREME SCORES, RANGES, MEANS AND STANDARD DEVIATIONS OF THE DISTRIBUTION OF SCORES MADE ON THE TEN VARIABLES EMPLOYED IN CANONICAL CORRELATION STUDY
(N = 46)

Variable Number	Variable	High Score	Low Score	Range	Mean	Standard Deviation
(3)	Professional Grade Point Average	3.00	1.27	1.73	2.33	00.4522
(4)	Major Teaching Field Grade Point Average	2.73	1.34	1.39	1.98	00.4058
(6)	Alphabets Test	3.35	2.14	1.21	2.52	00.2485
(7)	Two-digit Numbers Addition	108.00	34.00	74.00	67.40	13.7700
(8)	Two-digit Numbers Subtraction	90.00	29.00	61.00	58.54	13.1580
(9)	Two-digit Numbers Mixed	90.00	20.00	61.00	60.08	13.0010
(16)	General Activity	26.00	08.00	18.00	16.58	4.8040
(21)	Practice Teaching Grade	130.00	60.00	70.00	109.57	18.2900
(22)	Placement Bureau Rating	130.00	10.00	120.00	86.30	26.8900
(23)	Principal's M-Blank Rating	130.00	10.00	120.00	92.83	22.7100

ses of variables.

From the results of the analysis of variance, certain conclusions are warranted relative to good and average teachers as defined by ratings assigned by principals, Placement Bureau official and practice teaching supervisor:

1. There is very little difference between good teachers and average teachers in terms of emotional stability as measured by the Guilford-Zimmerman Temperament Survey.
2. Among the five personality traits measures, the greatest difference between good and average teachers is in terms of the General Activity personality trait. There is some difference between the two groups in terms of the Sociability and Ascendance traits as measured by the Guilford-Zimmerman Temperament Survey.
3. Good teachers in comparison to average teachers tend to differ in certain other respects: professional grade point average, major teaching field grade point average, flexibility in numerical abilities, and the disposition rigidity temperament factor as measured by the Alphabets test.

SECTION IV

ANALYSIS OF THE DATA—CANONICAL CORRELATION TREATMENT

THE ANALYSIS of the data in which canonical correlation method was employed will be presented in this section.

Following the analysis of group differences discussed in Section III, a further investigation of the data was made to determine the relationships between seven of the measures of teacher characteristics and a composite of the criterion ratings.

The seven measures employed in the canonical correlation analysis were selected from the original twenty measures on the basis of the profile differences found to exist between the group of good teachers and the group of average teachers. These profiles (Figure 1) show the amount of separation between the groups on each of the twenty measures. The seven variables on which the groups were separated the most were selected to be included in the study of relationships reported in this section.

These seven variables were: (1) professional grade point average; (2) major teaching field grade point average; (3) the Alphabets test (disposition rigidity measure); (4) Two-digit numbers, Addition; (5) Two-digit number, Subtraction; (6) Two-digit number, Mixed (flexibility in the numerical area); and (7) the General Activity personality trait.

The basic data for the canonical correlation analysis were the seven variables listed above and the three criterion estimates: Principal's M-Blank rating, practice teaching grade and Placement Bureau rating. The extreme scores, ranges, means and standard deviations of the distributions of scores on these ten variables are shown in Table X, preceding page.

A model of the canonical correlation method, given by Thomson (60), was followed in the study of relationships between the seven measures of teacher characteristics and a composite of the three criterion estimates.

Before presenting the canonical correlation analysis, a brief description of the method and a statement of the procedures followed will be given.

The canonical correlation is a multiple correlation between two sets of weighted variables. The classical multiple correlation is the situation in which a set of weighted variables is combined to predict only one criterion estimate. This criterion to be predicted in the classical multiple correlation may be one rating or a combination of ratings equally weighted to form one criterion score. The distinction between the classical multiple correlation and the canonical correlation is that in the solution for the canonical correlation, weights are found to apply to the several criterion measures to form a composite.

In brief, the procedures followed in this analysis were:

1. The weights for the criterion variables and the canonical correlation between the two weighted batteries were found by solving the determinantal equation given by Hotelling (32, 33).
2. Weights for the seven measures of teacher characteristics were found by using the usual regression procedures in which the seven measures were predicting the weighted composite of criterion ratings.

A preliminary step in the treatment of the data was to compute Pearson product moment coefficients of correlation as follows:

1. Among the three criterion variables;
2. Among the seven measures of teacher characteristics;
3. Between each of the seven measures and each criterion variable.

These correlations were then arranged in a ten by ten matrix. This matrix was partitioned and the various parts symbolized as shown in Figure 2.

TABLE XI

PARTITIONED CORRELATION MATRIX

	PT	PB	PM	PGPA	MGPA	Sub 7	TDN Add	TDN Sub	TDN Mix	G
PT	1.0	.504	.259	.451	.465	.016	.237	.133	.206	.150
PB	.504	1.0	.629	.500	.488	.236	.463	.353	.388	.436
PM	.259	.629	1.0	.293	.330	.368	.426	.403	.389	.470
PGPA	.451	.500	.293	1.0	.418	.078	.471	.389	.409	.381
MGPA	.465	.488	.330	.418	1.0	.208	.267	.206	.228	.157
Sub 7	.016	.236	.368	.078	.208	1.0	.005	.021	-.015	.262
TDN Add	.237	.463	.426	.471	.267	.005	1.0	.897	.878	.338
TDN Sub	.133	.353	.403	.389	.206	.021	.897	1.0	.900	.393
TDN Mix	.206	.388	.389	.409	.228	-.015	.878	.900	1.0	.313
G	.150	.436	.470	.381	.157	.262	.338	.393	.313	1.0

A	C^T
C	B

Figure 2. Symbols Used to Represent Sets of Correlations

The A matrix represented intercorrelations among the criterion variables. The B matrix represented intercorrelations among the seven measures of teacher characteristics. The C matrix represented correlations of each of the seven measures with each criterion variable. The C^T matrix was the transpose of C.

The partitioned matrix of actual correlations is presented in Table XI. These sets of correlations and intercorrelations, arranged as in Figure 2, were the basic data for the canonical correlation analysis.

The solution for the canonical correlation, as given by Hotelling (32, 33) is the equation

$$\left| C^T B^{-1} C - \lambda A \right| = 0$$

where C^T, B, C and A represent the matrices of intercorrelations shown in Table XI, and λ is the root which makes the equation have zero as its determinant. This root is the square of the canonical correlation, or R_c^2.

The inverse of B was computed, and after the appropriate multiplications the values in the matrix $C^T B^{-1} C$ were found. These two matrices are presented in Table XII and Table XIII.

The matrix $C^T B^{-1} C$ was substituted in the equation $C^T B^{-1} C - \lambda A = 0$ as follows:

$$\overset{C^T B^{-1} C}{\begin{vmatrix} .326642 & .332436 & .181567 \\ .332436 & .486724 & .391615 \\ .181567 & .391615 & .406062 \end{vmatrix}} - \overset{\lambda A}{\begin{vmatrix} \lambda 1.000 & \lambda 0.504 & \lambda 0.259 \\ \lambda 0.504 & \lambda 1.000 & \lambda 0.629 \\ \lambda 0.259 & \lambda 0.629 & \lambda 1.000 \end{vmatrix}} = 0$$

The value for λ, or the root for this determinantal equation, was found by trial and error, first taking for λ the square of the largest correlation among the correlations in C and working upward until the sign changed, then interpolating. Following this procedure,

Trial $\lambda = .615505$ gave

$$\begin{vmatrix} -.288863 & +.022222 & +.022151 \\ +.022222 & -.128781 & +.004462 \\ +.022151 & +.004462 & -.209443 \end{vmatrix} \qquad \overset{\text{(minus)}}{\triangle} = -.007614$$

Trial λ = .496505 gave

+.169363	+.082198	+.052972
+.082198	-.009781	+.079313
+.052972	+.079313	-.090443

(plus)
$\triangle$ = +.000411

Interpolating, λ was found to be .538456. The square root of this value was .734, which is the canonical correlation between the set of weighted criterion variables and the set of seven measures of teacher characteristics.

The standard error for this canonical correlation is $\frac{1}{\sqrt{N - 8 - 1}}$ or .17. Since the obtained correlation is greater than two and one-half times the standard error it can be concluded the true value for R_c is significantly different from zero (43).

Next, the weights for the criterion variables were found by solving equation $[C^TB^{-1}C - \lambda A]\begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0$ under the restriction that λ = .538456.

The values found by solving this equation with λ = .5384 were:

-.21184	+.061054	+.042107
+.061054	-.051732	+.052926
+.042107	+.052926	-.132394

$\triangle$ = -.00027216

The cofactors of these values were computed. The ratio in any of the three rows of cofactors were the weights for the criterion variables. The cofactors were:

+.00404785	+.01031174	+.00540962
+.01031174	+.02626990	+.01378127
+.00540962	+.01378127	+.00722997

The ratio in any of these rows of cofactors was: .3925 : 1.000 : .5246 . These values are the weights for the criterion variables.

Weights for the seven measures of teacher characteristics were found by the following sequence of operations: (1) condensing the A matrix using the weights found, (2) normalizing the vector now representing the set of criterion intercorrelations, and (3) performing the usual regression calculations with the seven variables predicting the normalized composite of the criterion variables.

The A matrix was condensed, using the calculated weights, as follows:

	.3925	1.0000	.5246							
.3925	1.0	.504	.259	.451	.465	.016	.237	.133	.206	.150
1.0000	.504	1.0	.629	.500	.488	.236	.463	.353	.388	.436
.5246	.259	.629	1.0	.293	.330	.368	.426	.403	.389	.470
	.451	.500	.293	1.0	.418	.078	.471	.389	.409	.381
	.465	.488	.330	.418	1.0	.208	.267	.206	.228	.157
	.016	.236	.368	.078	.208	1.0	.005	.021	-.015	.262
	.237	.463	.426	.471	.267	.005	1.0	.897	.878	.338
	.133	.353	.403	.389	.206	.021	.897	1.0	.900	.393
	.206	.388	.389	.409	.228	-.015	.878	.900	1.0	.313
	.150	.436	.470	.381	.157	.262	.338	.393	.313	1.0

The condensed values for the A, C and C^T matrices were:

2.5915	.8307	.8436	.4353	.7795	.6166	.6729	.7414
.8307	1.0	.418	.078	.471	.389	.409	.381
.8436	.418	1.0	.208	.267	.206	.228	.157
.4353	.078	.208	1.0	.005	.021	-.015	.262
.7795	.471	.267	.005	1.0	.897	.878	.338
.6166	.389	.206	.021	.897	1.0	.900	.393
.6729	.409	.228	-.015	.878	.900	1.0	.313
.7414	.381	.157	.262	.338	.393	.313	1.0

These condensed values were normalized, using the reciprocal of the square root of 2.5915 as a multiplier on each of the values. The new values were the correlation of the composite with itself and the correlations of each of the seven measures of teacher characteristics with the composite. The correlations of the seven measures with the composite criterion are shown in Table XIV.

The significance of these correlations was tested by the method given in McNemar (43). The standard error with three degrees of freedom was .1525. The obtained correlation was considered significantly different from zero if it was greater than two and one-half times this standard error.

The weights for the seven measures of teacher characteristics were found by post-multiplying B^{-1} by the normalized values in C. These weights are presented in Table XV.

Summary of Analysis of Data—Canonical Correlation Treatment

Relationships between a weighted composite of three criterion ratings and a weighted battery of seven measures of teacher characteristics were investigated by Hotelling's canonical correlation method. Results of this analysis were:

1. The canonical correlation between the weighted composite and the weighted battery of seven measures of teacher characteristics was .73. This correlation was significant at the 1% level.

When the criterion estimates are combined according to the weights found and when a linear combination of the seven measures is made using the weights found, it is possible to predict teaching success from these data in a limited time sense. Prediction from these data is limited because the measures of teacher characteristics employed in this study were obtained at the time the teachers were already on the job. The characteristics measured may have developed as a result of the teaching experiences and might not be present in persons who had not yet taught.

2. Weights for the three criterion estimates were:

Practice teaching grade	+0.392
Placement Bureau rating	+1.000
Placement M-Blank rating	+0.525

Combining these three ratings into a composite using these weights places greater emphasis on the Placement Bureau rating and Principal's M-Blank rating than on the Practice teaching grade.

3. Weights for the seven measures of teacher characteristics, and the correlation of each measure with the weighted composite criterion are presented in Table XVI.

Six of the seven measures of teacher characteristics are significantly correlated with the weighted composite criterion. Of these six, those which correlate highest with the composite are: Major teaching field grade point average; Professional grade point average; Two-digit number, Addition; and the General activity personality trait.

One objective test of temperament, the measure of flexibility in numerical abilities, is significantly correlated with the composite while the other objective test of temperament, a measure of disposition rigidity, is not. This measure of disposition rigidity had the lowest correlation with the composite.

SECTION V

SUMMARY AND CONCLUSIONS

Statement of the Problem

THE PRIMARY purpose of this study was to determine whether a group of good teachers could be differentiated from a group of average teachers in terms of certain measures of pre-service achievement, temperament and personality.

TABLE XIV

CORRELATION OF EACH OF THE SEVEN MEASURES OF TEACHER CHARACTERISTICS WITH THE COMPOSITE CRITERION

Variable	Correlation
Professional Grade Point Average	.516*
Major Teaching Field Grade Point Average	.524*
Reverse Strokes Test (Disposition Rigidity)	.270
Two-digit Numbers, Addition	.484*
Two-digit Numbers, Subtraction	.383*
Two-digit Numbers, Mixed (Flexibility)	.418*
General Activity Personality Trait	.461*

*Significant at the 1% level.

TABLE XV

WEIGHTS FOR EACH OF THE SEVEN MEASURES OF TEACHER CHARACTERISTICS

Measure of Teacher Characteristics	Weight
Professional Grade Point Average	+.1418
Major Teaching Field Grade Point Average	+.3714
Reverse Strokes (Disposition Rigidity)	+.1300
Two-digit Numbers, Addition	+.4619
Two-digit Numbers, Subtraction	-.4270
Two-digit Numbers, Mixed (Flexibility)	+.1815
General Activity Personality Trait	+.2776

TABLE XVI

SUMMARY OF WEIGHTS FOR SEVEN MEASURES OF TEACHER CHARACTERISTICS AND THE CORRELATION OF EACH MEASURE WITH THE WEIGHTED COMPOSITE CRITERION

Variable	Weights	Correlation with Composite
Professional Grade Point Average	+. 1418	. 516*
Major Teaching Field Grade Point Average	+3174	. 524*
Alphabets Test (Disposition Rigidity)	+. 1300	. 270
Two-digit Numbers, Addition	+. 4619	. 484*
Two-digit Numbers, Subtraction	-. 4270	. 383*
Two-digit Numbers, Mixed (Flexibility)	+. 1815	. 418*
General Activity vs., Inactivity	+. 2776	. 461*

*Significant at the 1% level.

A secondary purpose of the study was to investigate the relationships between a composite of three ratings of teaching success and certain measures of preservice achievement, temperament and personality.

Measures Employed in the Study

Twenty-three measures were employed in the study. These may be classified into four groups as follows:

1. Criterion estimates of teaching success
   (1) Practice teaching grade
   (2) Placement Bureau rating
   (3) Principal's M-Blank rating reported on Wisconsin Adaptation of M-Blank

2. Measures of preservice achievement. These measures were obtained from University of Wisconsin records. The four preservice measures were:
   (1) Quantitative score (expressed in Percentile Rank) American Council on Education Psychological Examination.
   (2) Linguistic score (expressed in Percentile Rank), American Council on Education Psychological Examination.
   (3) Professional grade point average earned in preservice education courses: Education 73, 74, and 75.
   (4) Grade point average earned in courses taken in the major teaching field.

3. Objective measures of temperament. These objective tests of temperament were speed tests designed to measure specific behavior of persons in some standard situation, rather than tests requiring self-evaluation on the part of the subject. Two kinds of objective tests were employed:
   (1) Creative effort tests developed by Cattell as measures of disposition rigidity
      a. Reverse Strokes test
      b. Alphabets test
   (2) Battery of alternation tests developed by Kleemeier and Dudek as measures of flexibility in three areas
      a. Numerical speed
      b. Verbal abilities
      c. Perceptual speed

4. Measures of personality. This group was composed of measures of five traits obtained by administering the Guilford-Zimmerman Temperament Survey. These five traits were:
   (1) General Activity vs. Inactivity, Slowness
   (2) Restraint vs. Impulsiveness
   (3) Ascendance vs. Submissiveness
   (4) Sociability vs. Shyness, Seclusiveness
   (5) Emotional Stability vs. Neuroticism

Design of the Study

Forty-six women teachers in their second, third or fourth year of teaching in Wisconsin secondary schools were the subjects for this study.

This group of forty-six teachers was divided into an upper half and a lower half on the basis of the three criterion estimates. The upper half was designated as Group I, good teachers, and the lower half was designated as Group II, average teachers.

Differences between these two groups in terms of certain measures of preservice achievement, temperament and personality were investigated by a multivariate analysis of variance technique.

Following this analysis of group differences, a study of relationships between a composite of the three criterion ratings and seven of the twenty measures of teacher characteristics was made by a canonical correlation method.

Results of the Study

There was a significant difference between the group of good teachers and the group of average teachers in terms of the profile of group means on the twenty measures of teacher characteristics. The profile differences showed the variables on which the groups differed the most were: (1) professional grade point average, (2) major teaching field grade point average, (3) flexibility in numerical abilities, (4) disposition rigidity as measured by the Alphabets test, (5) the General Activity personality trait.

There was no significant difference between the group of good teachers and the group of average teachers in terms of classes of variables (preservice achievement variables, temperament variables and personality variables).

A significant relationship (canonical correlation of .73) was found between a weighted composite of the three criterion estimates and a weighted combination of seven measures of preservice achievement, temperament and personality. In the weighted composite criterion, greater emphasis was placed on the Placement Bureau rating and Principal's M-Blank rating than on the practice teaching grade. In the combination of seven measures of teacher characteristics, greater emphasis was placed on the major teaching field grade point average; two-digit numbers, addition; and the General Activity personality trait.

Six of the seven measures of teacher characteristics employed in the canonical correlation analysis were significantly correlated with the composite criterion. These six measures were (1) professional grade point average, (2) major teaching field grade point average, (3) two-digit numbers, addition, (4) two-digit numbers, subtraction, (5) two-digit numbers, mixed, and (6) General Activity trait as measured by the Guilford-Zimmerman Temperament Survey. Of these six measures, those which correlated highest with the composite were the two grade point averages; General Activity personality trait; two-digit numbers, addition; and two-digit numbers, mixed (flexibility in the area of numerical abilities).

Conclusions

The significant differences found between the two groups of teachers seem to indicate that a pattern of measures of preservice achievement, temperament and personality will differentiate good teachers from average teachers as defined by the three criterion estimates employed in this study.

If one wishes to make interpretations from the objective measures of temperament and from the Guilford-Zimmerman Temperament Survey employed in this study, it is possible to compare good teachers with average teachers (as defined by the three criteria) in certain respects.

If these interpretations are made, good teachers would appear to be characterized as liking a rapid pace rather than a slow and deliberate one; they may be further characterized by a liking for quickness of action and production and efficiency. Good teachers appear to be higher academically in terms of both the professional and the major teaching field grade point averages than are average teachers. Good teachers would seem to be more flexible in numerical abilities than average teachers.

A relationship between teaching success as measured by a weighted composite criterion and a linear combination of measures of preservice achievement, temperament and personality seems indicated. This relationship probably cannot be considered predictive of teaching success except in a limited time sense. However, the correlations between the various measures and the weighted composite point out teacher characteristics which it might be profitable to investigate in further studies of the prediction of teaching success.

In further studies in this area, it would appear profitable to more thoroughly explore a greater number of temperament traits as measured by objective tests.

Further studies should be made along the same general lines, using different samples, to validate the relationships found in the present investigation.

BIBLIOGRAPHY

1. Allport, Gordon W. Personality, A Psychological Interpretation (New York: Henry Holt and Co., 1937).
2. Anderson, Harold M. A Study of Certain Criteria of Teaching Effectiveness, Ph. D. Thesis, University of Wisconsin, 1952.
3. Barr, A. S. Characteristic Differences in the Teaching Performance of Good and Poor Teachers of the Social Studies. (Bloomington, Ill.: Public School Publishing Co., 1929), 127 pp.
4. Barr, A. S. "The Measurement and Prediction of Teaching Efficiency: A Summary of Investigations," Journal of Experimental Education, XVI (June 1948), pp. 203-83.
5. Barr, A. S., and Emans, Lester M. "What Qualities are Prerequisite to Success?" Nation's Schools, VI (September 1930), pp. 60-64.
6. Baxter, B. Teacher-Pupil Relationships (New York: Macmillan Co., 1950).
7. Bimson, Oliver H. "Do Good Teachers Produce Good Results?" North Central Association Quarterly, XII (October 1937), pp. 271-76.
8. Block, J., and others. "Testing for Existence of Psychometric Patterns," Journal of Abnormal and Social Psychology (July 1951), pp. 356-359.
9. Buros, Oscar K., (ed.) Fourth Mental Measurements Yearbook (New Jersey: Gryphan Press, 1953).
10. Burt, Cyril. "Factor Analysis and Canonical Correlations," British Journal of Psychology, Statistical Section, I (July 1948), pp. 95-106.

10a. Cattell, Raymond B. Personality, A Systematic Theoretical and Factual Study (New York: McGraw-Hill Book Co., 1950).

11. Cattell, Raymond B. "The Riddle of Perseveration," Journal of Personality, XIV (1946), pp. 229-267.
12. Cattell, Raymond B. "Temperament Tests, II. Tests," British Journal of Psychology, XXIV (1933), pp. 20-49.
13. Cattell, Raymond B. A Guide to Mental Testing, 3rd edition (London: University of London Press, 1953).
14. Cattell, Raymond B., and Tiner, R. B. "The Varieties of Structural Rigidity," Journal of Personality, XVII (March 1949), pp. 321-341.

15. Cattell, Raymond B. "The Assessment of Teaching Ability," British Journal of Educational Psychology, Vol. I, Part 1 (February 1931), pp. 49-71.

16. Cattell, Raymond B. The Principal Invariant Personality Factors Established in Objective Tests, Advance Publication No. 1 (Urbana, Ill.: Laboratory of Personality Assessment and Group Behavior, May 1953).

17. Cattell, Raymond B. Description and Measurement of Personality (New York: World Book Co., 1946).

18. Cook, W. W., and Leeds, Carroll H. "Measuring the Teaching Personality," Educational and Psychological Measurement, VII (May 1947), pp. 399-410.

19. Cumbee, Frances, and Harris, Chester W. "The Composite Criterion and its Relation to Factor Analysis," Research Quarterly of the American Association for Health, Physical Education and Recreation, XXIV (May 1953), pp. 127-134.

20. Dixon, W. J., and Massey, F. J., Jr. Introduction to Statistical Analysis (New York: McGraw-Hill Book Co., 1951), pp. 290-294.

21. Domas, Simeon J., and Tiedeman, David V. "Teacher Competence: An Annotated Bibliography," Journal of Experimental Education, XIX (December 1950), pp. 101-218.

22. Edwards, Allen L. Experimental Design in Psychological Research (New York: Rinehart and Co., 1950), pp. 410-413.

23. Erickson, Harley E. A Factorial Study of Teaching Ability, Ph. D. Thesis, University of Wisconsin, 1952.

24. Evans, Kathleen M. "A Critical Survey of Methods of Assessing Teaching Ability," British Journal of Educational Psychology, XXI (June 1951), pp. 89-95.

25. Eysenck, Hans J. Dimensions of Personality (London: Broadway House, 1947), 308 pp.

26. Gotham, R. E. "Personality and Teaching Efficiency," Journal of Experimental Education, XIV (December 1945), pp. 157-165.

27. Guilford, J. P., and Guilford, R. B. "Personality Factors S, E, and M, and Their Measurements," Journal of Psychology, II (1936), pp. 107-127.

28. Guilford, J. P., and Guilford, R. B. "Personality Factors D, R, T, and A," Journal of Abnormal and Social Psychology, XXXIV (1939), pp. 21-36.

29. Guilford, J. P. Psychometric Methods, 2nd edition (New York: McGraw-Hill Book Co., 1954), 507 pp.

30. Guilford, J. P., and Zimmerman, W. S. Manual of Instructions and Interpretations (Beverly Hills, Calif.: Sheridan Supply Co., 1949).

31. Holzinger, Karl J. "Factoring Test Scores and Implications for the Method of Averages," Psychometrika, Vol. IX, No. 3 (September 1944), pp. 155-165.

32. Hotelling, Harold. "Relations Between Two Sets of Variates," Biometrika XXVIII (December 1936), pp. 321-377.

33. Hotelling, Harold. "The Most Predictable Criterion," Journal of Educational Psychology, XXVI (1935), pp. 139-142.

34. Humm, Doncaster C. "The Rationale of Temperament Testing," Educational and Psychological Measurement, V (1945), pp. 383-393.

35. Jones, Ronald D. "The Prediction of Teaching Efficiency from Objective Measures," Journal of Experimental Education, XV (September 1946), pp. 85-99.

36. Kemp, L. C. D. "The Personality of the Teacher," Forum of Education, Vol. V, No. 3 (May 1947), pp. 121-129.

37. Kinney, Lucian B. Measures of a Good Teacher (Stanford, California: Stanford University Press, 1953).

38. Kleemeier, R. W., and Dudek, R. J. "A Factorial Investigation of Flexibility," Educational and Psychological Measurement, Vol. X, No. 1 (Spring, 1950), pp. 107-118.

39. Kleemeier, Lyla, and Kleemeier, R. W. "Effects of Benzedrine Sulfate on Psychomatic Performance," American Journal of Psychology, LX (January 1947), pp. 89-100.

40. LaDuke, C. V. "The Measurement of Teaching Ability, Study No. 3," Journal of Experimental Education, XIV (September 1945), pp. 75-100.

41. Lamke, Tom A. Personality and Teaching Success, Ph. D. Thesis, University of Wisconsin, 1951.

41a. Lins, Leo J. "The Prediction of Teaching Efficiency," Journal of Experimental Education, XV (September 1946), pp. 2-60.

42. Lindquist, E. F. Educational Measurement (Washington, D. C.: American Council on Education, 1950), 819 pp.

43. McNemar, Quinn. Psychological Statistics (New York: John Wiley and Sons, 1950), 388 pp.

44. Menon, T. K. H., and Parekh, S. C. "Perseveration and Teaching Ability," Indian Journal of Psychology, XXV (1950), pp. 85-97.

45. Montross, Harold M. Temperament

and Teaching Success, Ph. D. Thesis, University of Wisconsin, 1952.

46. Mosier, C. L. "A Factor Analysis of Certain Neurotic Tendencies," Psychometrika, II (1937), pp. 263-287.

47. Notcutt, B. "Perseveration and Fluency," British Journal of Psychology, XXXIII (1943), pp. 200-208.

48. Phillips, Wendell S. An Analysis of Certain Characteristics of Active and Prospective Teachers, Contributions to Education, No. 161 (Nashville: George Peabody College for Teachers, 1935).

49. Pyle, W. H. "The Relation Between Intelligence and Teaching Success," Educational Administration and Supervision, XIV (April 1928), pp. 257-267.

50. "Report of the Committee on the Criteria of Teacher Effectiveness," Review of Educational Research, XXII (June 1952).

51. Rolfe, J. F. "The Measurement of Teaching Ability, Study No. 2," Journal of Experimental Education, XIV (September 1945), pp. 52-74.

52. Ryans, David G. "The Criteria of Teaching Effectiveness," Journal of Educational Research, XLII (May 1949), pp. 690-699.

53. Ryans, David G., and Wandt, Edwin. "Investigations of Personal and Social Characteristics of Teachers," Journal of Teacher Education, III (September 1952), pp. 228-231.

54. Schwartz, Anthony N. A Study of the Discriminating Efficiency of Certain Objective Tests of the Primary Source Personality Traits of Teachers, Ph. D. Thesis, University of Wisconsin, 1950.

55. Seagoe, May V. "Prediction of In-Service Success in Teaching," Journal of Educational Research, XXXIX (May 1946), pp. 658-663.

56. "Second Report of the Committee on Criteria of Teacher Effectiveness," Journal of Educational Research, XLVI (May 1953), pp. 642-658.

57. Spearman, Charles. The Abilities of Man (New York: Macmillan Co., 1927).

58. Stoelting, Gustave J. The Selection of Candidates for Teacher Education at the University of Wisconsin, Ph. D. Thesis, University of Wisconsin, 1954.

59. Tanner, William C. "Personality Bases in Teacher Selection," Phi Delta Kappan, XXXV (April 1954), pp. 271-277.

60. Thomson, Godfrey. "The Maximum Correlation of Two Weighted Batteries," British Journal of Psychology, Statistical Section, I (1947), pp. 27-34.

61. Walker, Helen, and Lev, Joseph. Statistical Inference (New York: Henry Holt and Co., 1953).

62. Walker, K. F., and others. "Perseveration Tests and the Concept of Mental Inertia," Character and Personality, XXI (1943), pp. 32-42.

Journal of Experimental Education

Volume XXV March, 1957 Number 3

COMPUTATIONAL ILLUSTRATIONS OF THE INTERNAL AND EXTERNAL CONSISTENCY ANALYSIS OF EXAMINATION RESPONSES

WILLIAM J. MOONAN*
USN Personnel Research Field Activity
San Diego, California

Introduction

THE FIRST purpose of this paper is to illustrate the computations involved in estimating the internal consistency of the responses to items of an examination which was given under several different experimental conditions with each condition utilizing different subjects. The second purpose is to show how to estimate the internal and external consistency of responses to the items of an examination which was administered to the same subjects at different times. The theoretical developments of these two problems were presented in references 5 and 6.

The coefficient of internal consistency measures the intra-class correlation of the responses to the items of an examination. If the responses possess very little internal consistency, then few individual differences can be detected among the subjects to whom the examination is given. Incidentally, the coefficient of internal consistency is a statistic which may be used in comparing two or more examinations without regard to the number of items of each. A coefficient of external consistency is also an intra-class correlation but it measures the consistency of the responses to the items or scores made by the same subjects in different administrations. An examination which is associated with a small coefficient of external consistency would be of limited value for identifying individual differences since the scores of an individual would vary nearly as much over the repeated administrations as would the scores of all individuals in any administration. In effect then, a score is not stable relative to the time or conditions of the administration of the examination and to the scores made by other individuals. Computational illustrations of the determination of coefficients of internal and external consistency are given in the following two sections.

Simultaneous Examination and Method Analysis (5)

The data for this illustration were taken from (4) which was concerned with evaluating the effects of a high school Solid Geometry prerequisite for students in college engineering drawing classes. There were three "methods" (M=3) involved. These are designated by the type of students who were enrolled in one of three groups. The three groups were:

Experimental Group—Students who entered the College of Engineering with a deficiency in high school Solid Geometry and who enrolled in a college course called Engineering Drawing - 4 prior to taking Solid Geometry in college.

Regular Group—Students who entered the College of Engineering without a deficiency in high school Solid Geometry and who enrolled in Engineering Drawing - 4.

Control Group—Students who entered the College of Engineering with a deficiency in high school Solid Geometry and who enrolled in Solid Geometry in college prior to enrolling in Engineering Drawing - 4.

A five-choice, multiple-choice examination consisting of fifty items (I = 50) was given to each of the individuals in each of the groups after they had completed Engineering Drawing - 4. The items of the examination concerned problems typically associated with orthographic projection. The responses to the items were scored dichotomously. Correct responses were scored as 1, otherwise they were scored 0. Total scores were equivalent to the total number of correct responses. W. G. Cochran (2) has shown that analysis of variance procedures are suitable for use with

Bureau Edhl. Psy. Research

---

*The opinions expressed are solely those of the author and are in no way official; nor are they to be construed as representing those of the U.S. Naval Personnel Research Field Activity or Bureau of Personnel.

data scored in this manner

The parametric form of a response to an item (i) by a subject (s) in one of the groups (m) is indicated by

$$y(ims) = \xi + \iota(i) + \mu(m) + \sigma(s) + \theta(im) + \varphi(is)$$

$$i = 1, \ldots, I; \quad m = 1, \ldots, M; \quad s = 1, \ldots, S(m)$$

where ξ represents the general effect, $\iota(i)$ the effect of the i^{th} item, $\mu(m)$ the effect of the m^{th} method, $\theta(im)$ the effect of the $(im)^{th}$ interaction and $\varphi(is)$ the effect of the $(is)^{th}$ interaction. It is further assumed that each response is associated with a variance, σ^2, and that the responses for a given subject are correlated to the degree $\rho(I)$ which is called the coefficient of internal consistency. If the examination consists of items which are not mutually intra-correlated to the same degree, the estimate of $\rho(I)$ will be a ratio of the average covariance and average variance. The analysis of variance of the y(ims) also assumes that all parameters involving an s parscript are random variables.

The data are summarized as sums and sums of squares (S. S.) in Part I of Table I. The quantities involved are designated according to the symbols used in Table I of (5). The layout, which is basic to the determination of the numerical values in Part I of Table I, will now be described. Consider, for each of the three groups, a rectangle of responses, which in this case are 1's and 0's, whose width corresponds to the number of items and whose length corresponds to the number of subjects in the group. If, for a given group, we total and record the 1's for individuals across the rows we obtain the total scores. If we total the 1's vertically down the columns we obtain the item totals. For each group we need to sum and find the sums of squares of the responses within the rectangle and to find the sum of squares of the item totals and total scores. For dichotomously scored (0, 1) items, the sums of squares of the responses is equal to their sum. Additionally, the totals of the same items of each group are added together and the sum of the squares of these totals (2046044) is found. The calculation formulas given in Part II of Table I are then evaluated for the data given in Part I. The summary of the analysis of variance is given at the top of Table II.

The 95% confidence interval (C. I.) of $\rho(I)$ which is a correlational property of the responses to the items and, also, the index of internal consistency, $\rho(H)$, will now be evaluated. The quantity $\rho(H)$ is a property of the entire examination since it is the Spearman-Brown function of the number of items, I, and $\rho(I)$, thus $\rho(H) = I\rho(I)/[1 + \overline{I-1}\rho(I)]$. A small $\rho(H)$ would indicate that compounding the responses of a subject to the I items of the examination does not greatly facilitate the detection of individual differences. The information given by the estimate of $\rho(H)$ is of the same type as that provided by "split-half" and other similar types of reliability coefficients if the Spearman-Brown formula has been used in conjunction with them. In these cases intraclass correlations should be used instead of product-moment correlations since the Spearman-Brown formula assumes that the variances of the subjects' scores are equal. The procedure given here has the advantage of being independent of any type of "split" made on the items to form the subtests. In some cases an unfortunate "split" may lead to a misleading index of internal consistency. To get the C. I.'s, we consult a variance ratio table and find the two tabulated values, $F[U] = F[345, 16905; .975] = 1.17$ and $F[L] = 1/F[16905, 345; .975] = 1/1.20 = .833$. From the variance column of Table II the statistic $F[o] = .9330/.1822 = 5.12$ is determined. Using equation 13 of (5) the interval is calculated to be

$$(1) \qquad \frac{5.12 - .833}{5.12 + 49(.833)} > \rho(I) > \frac{5.12 - 1.17}{5.12 + 49(1.17)}$$

$$.09 > \rho(I) > .06,$$

where $\overline{I-1} = 49$. Since the C. I. does not include zero, the hypothesis that $\rho(I) = 0$ (or $\rho(H) = 0$) is rejected at the .05 level of confidence. The point estimate of $\rho(I)$ is found by substituting 1.00 for 1.17 on the right side of (1). Thereby, $r(I) = .08$. The 95% C. I. for $\rho(H)$ is

$$(2) \qquad 1 - .833/5.12 > \rho(H) > 1 - 1.17/5.12$$

$$.84 > \rho(H) > .77$$

The point estimate of $\rho(H)$ is also found by substituting 1.00 for 1.17 but this is done now on the right side of (2). The quantity $r(H)$ is found to be .80.

The coefficient of internal consistency, $r(I) = .08$, for these data is quite low and an examination at least as long as 50 items of the same kind is needed if the examination is to be considered efficient for identifying individual differences. This is indicated by the moderate value $r(H) = .80$. The standard error of measurement (S. E. M.), $\sqrt{I\sigma^2[1 - \rho(I)]}$, expressed in terms of total score units is estimated by multiplying the variance of E[IS] by I = 50 and then taking the square root of that product. Thus the S. E. M. $= \sqrt{50(.1822)} = 3.02$. Also the standard deviation of the subjects' scores is found by multiplying the variance of E[S] by I = 50 and taking the square root of that product. The S. D. $= \sqrt{50(.9330)} = 6.83$. Separate evaluations of the r(I)'s, r(H)'s, S. E. M.'s and S. D.'s for each group have also been made and are presented at the bottom of Table II. It may be seen that they exhibit remarkable agree-

TABLE I

SUMS, SUMS OF SQUARES AND ANALYSIS OF VARIANCE CALCULATIONS FOR THE DATA TAKEN FROM REFERENCE (4)

I. Sums and Sums of Squares

Description	Symbol	Group			Total
		Experimental	Regular	Control	
Subjects	S(m)	S(1) = 51	S(2) = 274	S(3) = 23	S = 348
Sum of Responses	$\sum_{i}\sum_{s} y(ims)$	1216	7426	700	9342
S. S. of Responses	$\sum_{i}\sum_{s} y^2(ims)$	1216	7426	700	9342
S. S. of Item Totals	$\sum_{i}[\sum_{s} y(ims)]^2$	36390	1297838	11268	2046044*
S. S. of Total Scores	$\sum_{s}[\sum_{i} y(ims)]^2$	31460	213534	22658	267652

* This number corresponds to the numerator of A given below.

II. Analysis of Variance Calculations

$$A = \sum_{i}[\sum\sum_{ms} y(ims)]^2/S = 2046044/348 = 5879.4368$$

$$B = \sum_{m}\sum_{i}[\sum_{s} y(ims)]^2/S(m) = \frac{36390}{51} + \frac{1297838}{274} + \frac{11268}{23} = 5940.0775$$

$$C = \sum_{m}[\sum_{i}\sum_{s} y(ims)]^2/IS(m) = \frac{(1216)^2}{(50)(51)} + \frac{(7426)^2}{(50)(274)} + \frac{(700)^2}{(50)(23)} = 5031.1693$$

$$D = \sum_{m}\sum_{s}[\sum_{i} y(ims)]^2/I = 267652/50 = 5353.0400$$

$$G = [\sum\sum\sum_{ims} y(ims)]^2/IS = (9342)^2/(50)(348) = 5015.6876$$

$$T = \sum\sum\sum_{ims} y^2(ims) = 7426 + 1216 + 700 = 9342.0000$$

TABLE II

ANALYSIS OF VARIANCE OF THE y(ims) AND SUMMARY STATISTICS BY GROUPS FOR THE DATA TAKEN FROM REFERENCE (4)

I. Analysis of Variance of the y(ims)

Source of Variation	Degrees of Freedom	Sum of Squares	Variance	F
$H[I]; \iota(i) = 0$	$\overline{I-1} = 49$	A-G = 863.7492	17.6275	96.75**
$H[IM]; \theta(im) = 0$	$\overline{I-1}\,\overline{M-1} = 98$	B-A-C+G = 45.1590	.4608	2.53**
$E[IS]$	$\overline{I-1}\,\overline{S-M} = 16905$	T-B-D+C = 3080.0518	.1822	
Intra-Subject Total	$\overline{I-1}\ S = 17052$	T-D = 3988.9600		
$H[o]; \xi = 0$	1	G = 5015.6876		
$H[M]; \mu(m) = 0$	$\overline{M-1} = 2$	C-G = 15.4817	7.7409	8.30**
$E[S]$	$\overline{S-M} = 345$	D-C = 321.8707	.9330	
Inter-Subject Total	S = 348	D = 5353.0400		
Grand Total	IS = 17400	T = 9342.0000		

** Hypothesis is rejected at least at the .01 level of significance.

II. Summary Statistics by Groups

Statistic	Group			Composite
	Experimental	Regular	Control	
Items	50	50	50	50
Subjects	51	274	23	348
Means	23.84	27.10	30.43	26.84
r(I)	.08	.07	.11	.08
r(H)	.81	.80	.86	.80
S.E.M.	3.04	3.02	2.91	3.02
S.D. (Scores)	7.02	6.71	7.84	6.83
Mean Item Diff.	.48	.54	.61	.54
S.D. (Item Diff.)	.23	.23	.24	.23

ment from group to group.

It is interesting to note from Table II that there exists some interaction between the items and the groups (F = 2.53**). This means that some items are more difficult for some groups than for others. This tends to complicate a straightforward interpretation of the item and method effects. The identification of the interacting items will not be undertaken here. However, scrutiny of the data leads to a judgment that the interaction effects are small compared to the amounts of group differences. There seems to be little doubt that a difference in the difficulties of the items has been detected, (F = 96.75**), and that the average difficulty, 26.84/50 = .54, seems quite satisfactory although the standard deviation of the difficulties, $\sqrt{17.6275/348}$ = .23, is quite large. This same standard deviation prevails approximately for all groups although the average difficulties vary according to the group means (.48, .54, and .61). The difficulty values of the items for each group are quite variable although the average difficulty is about .50 which is a value often recommended for psychometric examinations.

The hypothesis of equality of the group effects, H[M]; μ(m) = 0, was rejected, (F = 8.30**). The group means are given at the bottom of Table II. Duncan's multiple range test (α = .05) shows that each of the group means differs significantly from every other group mean. According to the author of (4) the approximate three point and six point differentials which exist between the group means represents a practical as well as a statistical difference. The evidence thus indicates that having had a course in Solid Geometry is beneficial to those taking Engineering Drawing—4.

The Determination of External and Internal Consistency (6)

The illustrations of the calculations which would be involved in the evaluation of the coefficients of external and internal consistency, according to (6) will now be described. For the example, data collected for the purpose of developing and scoring a proficiency test for a Tactical Range Recorder (TRR) operator (1) will be analyzed. The TRR is an instrument used, with other equipments, in making anti-submarine attacks and is part of the SONAR gear on U. S. Navy ships.

The examination analyzed in this example consisted of twenty-five items (I = 25) which were associated with the determination of "Range". Range is a measure of the distance a submarine is located from the SONAR gear and is one, among other kinds of information indicated by the TRR. The proficiency examination consisted of recorded signals of submarines operating under sea conditions. These signals were recorded on phono-discs and used to activate twelve TRR trainers which were operated by 104 subjects (S = 104) who were learning to become sonarmen. The task of the student was to evaluate the range for each of the 25 signals (items). For reasons given in (1) the responses were scored dichotomously.* A score of 1 was given to the response if it did not differ by a given amount from the true distance between the submarine and the SONAR gear. Otherwise the response was scored 0. The examination was administered twice (A = 2) during the training period of the students. These administrations occurred during the 16th and 21st weeks of training and will be so designated.

The linear model of a response to an item (i) in an administration (a) by one of the subjects (s) may be written as:

$$y(ias) = \mu + \iota(i) + \alpha(a) + \sigma(s) + \varphi(ia) + \theta(is) + \psi(as) + \varepsilon(ias)$$

$$i=1, \ldots, 25=I; \quad a=1, \ldots, 2=A; \quad s=1, \ldots, 104=S.$$

The parameters are: μ the general effect, ι(i) the effect of the i^{th} item, α(a) the effect of the a^{th} administration, σ(s) the effect of the s^{th} student and all two and three factor interactions defined by their parscripts. For instance φ(ia) is the effect of the interaction between the i^{th} item and the a^{th} administration. The variance of a response is σ^2 and it is assumed that responses by the same subject within an administration are intra-correlated to the degree ρ(I) which is referred to as the coefficient of internal consistency. Responses by a subject to the same item in different administrations are intra-correlated to the degree ρ(A) which is referred to as the coefficient of external consistency. In the development of the analysis of variance of the y(ias) it was also assumed that all parameters involving an s parscript are random variables and that responses by the same subject to different items in different administrations were correlated to the degree ρ = ρ(A) ρ(I). It seems reasonable to suppose that the intra-correlation of responses to different items in different administrations should be less than the intra-correlation of different items in the same administration and less than the intra-correlation of the same items in different administrations. Therefore, according to the logic of the problem, ρ should be less than either $\rho(A)$ or $\rho(I)$.

---

*Although both examples in this paper utilized dichotomously scored responses, the analyses made here do not depend upon this method of scoring.

Different choices of ρ lead to different forms of the analysis of variance of the y(ias). In general, the choice $\rho = \rho(A)\,\rho(I)$ seems most appropriate for the evaluation of responses to items in different administrations. However, this need not necessarily be the case for all types of items and conditions of administration. Thus, the analysis for a general value of ρ needs to be considered. The responses, y(ias), have been analyzed in (6) in the form of a 3-way classification analysis of variance. The expected variance, $\mathcal{E}$ (Var), of the E[IAS], E[IS], E[AS], and I[S] sources of variation, for a general value of ρ, are given below:

$$\begin{aligned}
\mathcal{E}\,(\text{Var.}[IAS]) &= \sigma^2[1-\rho(I)-\rho(A)+\rho].\\
\mathcal{E}\,(\text{Var.}[IS]) &= \sigma^2[1-\rho(I)+\overline{A-1}\,\rho(A)-\overline{A-1}\,\rho].\\
\mathcal{E}\,(\text{Var.}[AS]) &= \sigma^2[1-\rho(A)+\overline{I-1}\,\rho(I)-\overline{I-1}\,\rho].\\
\mathcal{E}\,(\text{Var.}[S]) &= \sigma^2[1+\overline{I-1}\,\rho(I)+\overline{A-1}\,\rho(A)+\overline{A-1}\;\overline{I-1}\,\rho].
\end{aligned}$$

The expected mean squares given in (6) agree with those given above if ρ is set equal to $\rho(A)\,\rho(I)$. In the event that a ρ other than $\rho(A)\,\rho(I)$ is assumed, a different method than the one shown here or in (6) must be used to estimate $\rho(A)$, $\rho(I)$ and ρ.

The choice $\rho = \rho(A)$, is suitable for the analysis of data arising from certain experimental conditions. R. O. Collier (3) develops this model theoretically, as well as many other analysis of variance models which involve correlated observations. A numerical illustration is also given in (3) for which the assumption $\rho = \rho(A)$ is appropriate. However, the illustration is not concerned with the type of psychometric analysis used here.

The sums and sums of squares which are basic to the analysis of variance of the responses are given in Table III. The literal values A, B, etc., of Table III are defined as in Table IV of (6). The layout of the responses is similar to that used in the previous example. The response rectangle is again laid out for each administration, and the sums of squares of the individual responses within the administrations, are calculated. Finally, in effect, the rectangles are superimposed on each other and the respective values summed. Then the sums (3385), and sums of squares of the totaled responses (5839), item totals (485803), and score totals (115915), are obtained. The analysis of variance is presented at the top of Table IV.

In order to determine the 95% C.I. of the $\rho(I)$ and $\rho(H)$ a variance ratio table is used to find $F[IU] = F[103, 2472; .975] = 1.31$, and $F[IL] = 1/F[2472, 103; .975] = 1/1.36 = .735$. Using information from the variance column of Table IV, $F[I] = 1.1145/.1898 = 5.87$, and the interval for $\rho(I)$ is:

$$(3) \qquad \frac{5.87 - .735}{5.87 + 24(.735)} > \rho(I) > \frac{5.87 - 1.31}{5.87 + 24(1.31)}$$

$$.22 > \rho(I) > .12,$$

where $24 = \overline{I-1}$. The confidence interval does not include zero so the hypothesis that $\rho(I) = 0$ (or $\rho(H) = 0$) is rejected at the .05 level of significance. The point estimate of $\rho(I)$ is found by substituting 1.00 for 1.31 on the right side of (3). Thus $r(I) = .16$. The index of internal consistency, $\rho(H)$, has a 95% C.I. given by

$$(4) \qquad 1 - .735/5.87 > \rho(H) > 1 - 1.31/5.87$$

$$.87 > \rho(H) > .77$$

with the value of $r(H) = .83$ when we replace 1.31 by 1.00 in (4).

To obtain a 95% C.I. for $\rho(A)$, the coefficient of external consistency, we determine $F[AU] = F[103, 103; .975] = 1.48$, $F[AL] = 1/F[103, 103; .975] = 1/1.48 = .676$, and $F[A] = 1.1145/.8941 = 1.25$. In this special case where $A = 2$, $F[AL] = 1/F[AU]$. With these numbers the interval is calculated as

$$(5) \qquad \frac{1.25 - .676}{1.25 + 1(.676)} > \rho(A) > \frac{1.25 - 1.48}{1.25 + 1(1.48)}$$

$$.30 > \rho(A) > -.08$$

where $\overline{A-1} = 1$. This confidence interval includes zero so that we accept the hypothesis that $\rho(A) = 0$ at the .05 level of significance. The point estimate of $\rho(A)$ is found by substituting 1.00 for 1.48 on the right of (5). The quantity $r(A) = .11$ is quite small. The responses to the twenty-five items evidently produce scores which are relatively inconsistent between the 16th and 21st weeks of training. It was indicated in (6) that for the special case of $A = 2$, which is true for this problem, the product-moment correlation of the total scores in each administration would be arithmetically equal or larger than $r(A)$. This correlation was evaluated to be .11 which, to the accuracy carried, is identical with $r(A)$. Several factors may account for the low coefficient of external consistency. One of them may be the scoring system used and another, particularly emphasized in (1), is the variance ascribed to the TRR equipment which was used in the study. The equipment effects were not estimated although an attempt was made to calibrate the instruments.

It is also possible to obtain an index of external consistency from the analysis of variance table. This index was not described in (6) but is related to $\rho(A)$ by the following formula:

TABLE III

SUMS, SUMS OF SQUARES AND ANALYSIS OF VARIANCE CALCULATIONS FOR THE DATA TAKEN FROM REFERENCE (1)

I. Sums and Sums of Squares

Description	Symbol	Administration		Total
		16th Week	21st Week	
Sum of Responses	$\underset{is}{\sum\sum} y(ias)$	1651	1734	3385
S. S. of Responses	$\underset{is}{\sum\sum} y^2(ias)$	1651	1734	5839*
S. S. of Item Totals	$\sum_i[\sum_s y(ias)]^2$	117095	126406	485803*
S. S. of Total Scores	$\sum_s[\sum_i y(ias)]^2$	28467	31826	115915*

* These numbers correspond to the numerators of E, A, and C, respectively.

II. Analysis of Variance Calculations

$$A = \sum_i[\underset{as}{\sum\sum} y(ias)]^2/AS = 485803/(2)(104) = 2335.5913$$

$$B = \sum_a[\underset{is}{\sum\sum} y(ias)]^2/IS = [(1651)^2 + (1734)^2]/(25)(104) = 2204.8296$$

$$C = \sum_s[\underset{ia}{\sum\sum} y(ias)]^2/IA = 115915/(25)(2) = 2318.3000$$

$$D = \underset{ia}{\sum\sum}[\sum_s y(ias)]^2/S = [117095 + 126406]/104 = 2341.3558$$

$$E = \underset{is}{\sum\sum}[\sum_a y(ias)]^2/A = 5839/2 = 2919.5000$$

$$F = \underset{as}{\sum\sum}[\sum_i y(ias)]^2/I = [28467 + 31826]/25 = 2411.7200$$

$$G = [\underset{ias}{\sum\sum\sum} y(ias)]^2/IAS = (3385)^2/(25)(2)(104) = 2203.5048$$

$$T = \underset{ias}{\sum\sum\sum} y^2(ias) = 1651 + 1734 = 3385.0000$$

TABLE IV

ANALYSIS OF VARIANCE OF THE y(ias) AND SUMMARY STATISTICS BY WEEKS OF TRAINING FOR THE DATA TAKEN FROM REFERENCE (1)

I. Analysis of Variance of the y(ias)

Source of Variation	Degrees of Freedom	Sum of Squares	Variance	F
H[IA]; $\varphi(ia) = 0$	$\overline{I-1}\ \overline{A-1} = 24$	D-A-B+G = 4.4397	.1850	1.24
E[IAS]	$\overline{I-1}\ \overline{A-1}\ \overline{S-1} = 2472$	T-D-E-F+A+B+C-G = 367.6403	.1487	
H[I]; $\iota(i) = 0$	$\overline{I-1} = 24$	A-G = 132.0865	5.5036	29.00**
E[IS]	$\overline{I-1}\ \overline{S-1} = 2472$	E-A-C+G = 469.1135	.1898	
H[A]; $\alpha(a) = 0$	$\overline{A-1} = 1$	B-G = 1.3248	1.3248	1.48
E[AS]	$\overline{A-1}\ \overline{S-1} = 103$	F-B-C+G = 92.0952	.8941	
H[M]; $\mu = 0$	1	G = 2203.5048	......	
E[S]	$\overline{S-1} = 103$	C-G = 114.7952	1.1145	
Total	IAS = 5200	T = 3385.0000		

** Hypothesis rejected at least at the .01 level of significance.

II. Summary Statistics by Weeks of Training

Statistic	16th Week	21st Week	Composite
Items	25	25	25
Subjects	104	104	208
Mean Score	15.88	16.67	16.27
r(I)	.14	.19	.16
r(H)	.80	.86	.82
r(A)	...	...	.11
r(B)	...	...	.20
S. E. M.	2.10	2.02	6.69
S. D. (Scores)	4.68	5.32	7.46
Mean (Item Diff.)	.64	.67	.65
S. D. (Item Diff.)	.18	.15	.16

$$(6) \qquad \rho(B) = \frac{A\,\rho(A)}{1+\overline{A-1}\,\rho(A)}$$

The relationship of ρ(A) to ρ(B) corresponds to the relationship of ρ(I) to ρ(H). The quantity r(B), which estimates ρ(B), would be useful in evaluating scores combined over the administrations for each subject. A large r(B) would indicate that compounding the scores over the administrations is effective for detecting individual differences between the subjects.

The 95% C.I. for ρ(B) is given by the wide interval

$$(7)\ 1 - .676/1.25 > \rho(B) > 1 - 1.48/1.25$$

$$.46 > \rho(B) > -.18$$

The point estimate of r(B) is obtained from the right side of (7) by replacing 1.48 by 1.00. In this case r(B) = .20 which agrees with the sample estimate form of (6), $r(B) = 2(.11)/[1+1(.11)] = .22/1.11 = .20$. The estimate of r(B) is low and the hypothesis that ρ(B) = 0 is accepted at the .05 level of significance. Thus the composite scores are of very limited value for detecting individual differences.

The analysis of variance table shows that the hypothesis of no administration effects with a one-sided alternative hypothesis was accepted (F = 1.48). The estimated average gain of the students over the five week training period was only about 1 test point, 16.67 - 15.88 = .79. This fact indicates that on the average, learning during the period was meager for the students. Table IV also indicates that there was little item-administration interaction, (F = 1.24), and that the items differed considerably in difficulty (F = 29.00**). The average difficulty in the 16th week was 15.88/25 = .64 and in the 21st week it was 16.67/25 = .67. The composite average difficulty was 16.27/25 = .65. These numbers reveal that, on the average, the items were relatively easy for the students, since the average difficulty is greater than .50. Tightening the tolerance used in scoring the responses would remedy this. The standard deviation of the composite difficulties was $\sqrt{5.5036/208} = .16$. For the separate administrations the item difficulty standard deviations were .18 and .15.

The standard error of measurement $\sqrt{I\sigma^2[1 - \rho(I)]}$, for a total score obtained on a single administration may be found by noting that the expected value of the variance of E[IS] is $\sigma^2[1 - \rho(I)][1 + \overline{A-1}\,\rho(A)]$. Since the estimated value of ρ(A) is .11, $\overline{A-1} = 1$, and the variance of E[IS] from Table IV is .1898, the quantity .1898/1.11 = .1710 is computed. To obtain this standard error of measurement (S.E.M.), .1710 is multipled by the number of items, I = 25, and the square root of the resulting number is taken. Thus, $\sqrt{25(.1710)} = 2.07$. The S.E.M.'s as estimated from data from each of the separate administrations is shown in Part II of Table IV. These composite scores, that is, the total score obtained from adding the separate totals obtained by each subject on the two administrations also has a standard error of measurement. To find this number, the variance of E[AS] is multiplied by IA = (25)(2), and the square root is taken. Therefore $\sqrt{(25)(2)(.8941)} = 6.69$. This number is large relative to the standard deviation of the composite scores, $\sqrt{(25(2)(1.1145)} = 7.46$, and indicates at once that the estimate of ρ(A) should be small. The standard deviation of scores between administrations is almost as large as the standard deviation of the composite scores. The estimate of the standard deviation of the total score of a single administration is found by dividing the variance of E[S] by $1 + \overline{A-1}\,r(A) = 1.11$, multiplying this quotient by I = 25 and taking the square root. Thus, the S.D. = $\sqrt{(25)(1.1145)/1.11}=5.01$. The S.D.'s which were estimated separately from data of each of the administrations are shown at the bottom of Table IV.

Estimates of ρ(I), ρ(H), ρ(A), and ρ(B) may be obtained by using another F[I] and F[A] calculated from the variance column of Table IV. When F[I] = .8941/.1487 = 6.01, r(I) = .17, r(H) = .83 and these numbers are very like the r(I) = .16 and r(H) = .82, which were previously obtained. Also F[A] can be evaluated as .1898/ .1487 = 1.28, and this gives r(A) = .12 and r(B) = .22, which are close to the other r(A) = .11 and r(B) = .20.

In lieu of the evidence presented, it appears that an examination consisting of the 25 range items which is administered to SONAR students during their 16th and 21st weeks of training, permits the detection of individual differences to a moderate degree, r(H) = .82, on a single administration. The coefficient of internal consistency, r(I) = .16, although numerically small, actually is quite large in comparison to other examination data which have been analyzed by the method. The consistency of responses to the same items, or equivalently the consistency of the scores between the two administrations is, however, very low, r(A) = .11, and this value was not statistically significant at the .05 level of significance. The two sets of responses are internally consistent to a favorable degree within sets. They are unfavorably externally consistent. This condition means that although individual differences can be detected to a moderate degree, for a given administration because of the length of the examination, the nature of these differences is not consistent over the time period of five weeks. Of course the analysis cannot account for this unusual circumstance, but further research needs to be undertaken to determine possible causes for it. Repeating this analysis for another sample under the same conditions has not been possible as yet. When this can be done, perhaps the scores for the 18th week of training can be collected and the effects of the TRR's used in the study can be controlled.

Totaling the scores for each subject over the two administrations does not facilitate the detection of individual differences since the scores made on each administration, being almost externally inconsistent, will total to a set which have a variance not much greater than the standard error of measurement which would be used to detect individual differences in the composite scores. This example is unique in that it shows that it is possible to obtain high internal consistency and low external consistency of responses to items of homogeneous material from subjects who take the same examination at two different times under similar conditions. Admittedly, the administrations were quite far apart in respect to the usual test-retest situtation, but nevertheless, the results were quite unusual.

Summary

The arithmetic operations involved in estimating coefficients of external and internal consistency of the responses to items from an examination have been illustrated. Two kinds of analyses were involved. One concerned the evaluation of the internal consistency of the responses in a single experimental situation wherein three methods (groups), were being compared. This corresponded to an analysis given in (5), and used data taken from (4). The other analysis concerned the evaluation of both internal and external consistency coefficients and indices from responses given by students to the items on an examination which was administered on two different occasions. Reference (6) provided the basis for the analysis and (1) supplied the data.

The analyses of variance which were used in these two examples provide an efficient and convenient means of determining the usual statistics necessary to make a psychometric analysis. The coefficients and indices of external and internal consistency are estimated easily by point and by interval. Also the standard deviations of the scores, the standard errors of measurement and certain item statistics are readily calculated. Additionally, tests of hypotheses regarding the parameters of the response model are easily made.

REFERENCES

1. Batterton, R. L. The Development of a Proficiency Test for the TRR Operator, Report No. 80 (San Diego, California: USN Personnel Research Field Activity, 1955).
2. Cochran, W. G. "The Comparison of Percentages in Matched Samples," Biometrika, XXXVII (1950), p. 256.
3. Collier, R. O. Experimental Designs in Which the Observations Are Assumed to be Correlated, Unpublished Ph. D. Dissertation, University of Minnesota, Minneapolis, Minnesota, 1956.
4. Clausen, J. A. The Effects of an Entrance Requirement on Selected Engineering Subjects, Unpublished Ph. D. Dissertation, University of Minnesota, Minneapolis, Minnesota, 1955.
5. Moonan, W. J. "Simultaneous Examination and Method Analysis by Variance Algebra," Journal of Experimental Education, XXIII (1955), p. 253.
6. Moonan, W. J. "An Analysis of Variance Method for Determining the External and Internal Consistency of an Examination," Journal of Experimental Education, XXIV (1956), p. 239.

THE RELATIVE ACHIEVEMENT OF THE OBJECTIVES OF ELEMENTARY SCHOOL SCIENCE IN MINNESOTA SCHOOLS

JANE JOHNSTON
Moorhead State Teachers College
Moorhead, Minnesota

TO WHAT extent are the objectives of elementary school science being achieved in Minnesota schools? What pupil, teacher, and teaching-situation factors contribute to the achievement of these objectives? What are the implications for the education of elementary school teachers? These problems were considered in a recent analytical survey of elementary school science in Minnesota.[1]*

For various parts of the survey the sampling unit was the school system (used only in securing preliminary data), the fifth-grade teacher, or the fifth-grade class (a cluster sample).

A brief preliminary questionnaire dealing with elementary school science teaching facilities and procedures was sent to 478 superintendents of public graded elementary school systems in Minnesota, 1952.

A proportionate stratified random sampling of 87 fifth-grade teachers from the same schools was secured in the following manner:

1. Schools were divided into four strata according to the number of teachers in their elementary schools—over fifty teachers (Group I), twenty-one through fifty teachers (Group II), eleven through 20 teachers (Group III), and one through ten teachers (Group IV).
2. Group I, the smallest group, contained eleven schools. It was decided to include two schools (18.2 percent) from this group, and the same percentage of schools from each of the other three groups.
3. The names of the school systems were then listed alphabetically for each stratum, and were numbered in that order. Appropriate random number tables were used to select from this listing two schools from Group I, seven from Group II, twelve from Group III, and sixty-six from Group IV.

A twenty-two item questionnaire dealing with teaching experience, colleges attended and courses taken, materials and facilities available for elementary school science teaching, methods used in teaching elementary school science, and time and emphasis placed on science in the classroom was sent to the eighty-seven fifth-grade teachers participating in the survey.

The substantial agreement between answers to similar questions on the superintendents' and teachers' questionnaires, and the fact that complete questionnaire returns were secured from over 98 percent of the superintendents and 95 percent of the teachers, provided checks on the reliability and completeness of the information provided.

A log of science activities in their classrooms was kept by a sub-sample of thirty of the eighty-seven participating teachers. The teachers in this sample were selected by random sampling methods (random numbers) and were assigned at random to three different time periods of ten days each. Information from these logs provided a check on information in the two questionnaires, and furnished additional information on the science activities and teaching methods that were being used.

According to questionnaire and log replies (1954), emphasis on science in elementary classrooms was less than that given to social studies or reading, and more than that given to music or art; the typical science class was thirty minutes long, and the average time spent on science per week was under two hours. About sixty percent of the schools had some purchased science equipment. Although most of the teachers reported that equal emphasis was given to biological and physical science, their logs indicated that biological science topics were considered in a ratio of 3:1 to physical science topics. Text reading and discussion was the "most-used" science teaching method, while field trips and laboratory work were the "least-used." The typical teacher had taken college courses in biology and health, had not taken courses in physical science or the teaching of science, and did not have a college degree.

Information on fifth-grade pupils taught by the eighty-seven participating teachers was secured by means of a science test (administered in December 1953, and again in May 1954), an intelligence test, and a science reading test. One hundred percent returns were obtained.

The 103-item science test, constructed by the writer, was designed to test knowledge of science facts, relationships, and generalizations.

Ednl Res
DAVID HA
Dated
No

---

*All footnotes will be found at end of article.

Subject matter included in the test was determined by (a) examining the science section of the Minnesota Curriculum Bulletin Number 7, A Guide for Teaching Science and Conservation[2], and including some questions on each phase of science mentioned; (b) including additional biological science questions, because that section of the Curriculum Bulletin is limited; (c) omitting questions on health, conservation, and safety because these topics have traditionally been included in social studies units, and it was desired to find out how much science the pupils were learning; and (d) including some questions from each of the important areas of science—as determined from examination of children's textbooks, curriculum bulletins, textbooks on science teaching, and research recommendations—such areas as plants and animals, the various forms of energy, transportation and communication, simple chemistry, astronomy, earth science, simple machines, and size and time relationships.

Before its administration, the test was submitted for criticism to six expert teachers of science, and a trial form of the test was administered to two fifth-grade and two sixth-grade classes, in order to detect confusing directions or questions, and to estimate the time needed to administer the test. Teachers who administered the first tests in December did not know that the same test was to be repeated the following May, and all tests were returned to the writer after the first administration to the eighty-seven fifth-grade classes. Test results on both the pretest and retest were secured from all eighty-seven of the participating fifth grades. The reliability of the science test was calculated by the split-test method, obtaining the maximum likelihood estimate.[3] The reliability of the pretest was found to be .91, and of the retest, .94.

Statistical techniques used in analyzing test and questionnaire results included the following:

1. The "t" test was used to determine the significance of the differences between means on the science pretest and retest.[4]

2. Cluster sample statistical treatment was used in determining the sample estimate of the population mean and variance of the mean on the science tests. In the testing program, the fifth-grade class to be tested was selected by random sampling methods. Thus each fifth-grade class in the population stratum had an equal chance of being selected. This fifth-grade class was, however, a cluster sample; each individual pupil in the whole population stratum of fifth-grade pupils in Minnesota did not have an equal chance of being selected. In calculating the estimated mean and variance for the population from this sample, therefore, cluster sample statistical analysis was used, in which each mean was weighted by the size of that particular fifth-grade class. The calculation formula for the population mean was:

$$\overline{X}_c = \frac{\sum_{i=1}^{m} n_i\overline{X}_i}{\sum_{i=1}^{m} n_i} = \frac{\Sigma\Sigma X_i}{N}$$

where $n_i\overline{X}_i$ is the number of individuals in the i^{th} cluster multiplied by the mean of the i^{th} cluster, and the estimated population mean is the sum of all of these quantities calculated for each of the eighty-seven groups and divided by the total number of individuals in the sample.[5]

Results of the above calculations showed the sample estimate of the population mean to be 59.54 for the science pretest and 64.67 for the retest, with a mean gain of 5.13 points. When the "t" test of significance was applied to this gain, it was found to be significant at the .01 level.

3. The analysis of variance statistical treatment was used to compare science test scores received by pupils in three different I.Q. groups. The following results were found: (a) Of the total group of 1850 pupils, 147 were found to have I.Q.'s of 121 plus; 1153 to have I.Q.'s of 95-115; and 207 to have I.Q.'s of below 90. Science test sub-score means on physical science and generalization questions were found to be significantly different (.01 level) for the three I.Q. groups. (b) Mean gains of the three I.Q. groups between the pretest and retest were not found to be significantly different.

4. The analysis of variance and analysis of variance and covariance statistical techniques were used to relate mean gains of fifth-grade classes between the science pretest and retest to each of the following teacher or teaching-situation factors: teachers' years in college; number of science areas in which teacher took college courses; teachers' years of teaching experience; whether or not the teacher took a college course in science teaching methods; college physical science courses taken by the teacher; use of science instruments for observing and measuring; number of science books in the classroom other than textbooks; time and emphasis placed on science; and the relative use of each of the following in science teaching:

Visual aids
Group planning of science projects
Group performing of science projects
Research reading
Teacher-demonstration and explanation
Individual laboratory work

A statistically significant relationship between the teacher-factor being considered and class gains on the science tests was found in one

TABLE I

ANALYSIS OF VARIANCE AND COVARIANCE, CHANGE BETWEEN SCIENCE PRETEST-RETEST AND TEACHER'S TEACHING EXPERIENCE, WITH PRETEST CONSTANT

Code	N	ΣXY	Σxy
1, 2	28	6,535.4358	- 80.0191
3, 4	36	12,517.0500	-295.1110
5	19	6,728.0346	-113.8774
Total	83	25,780.5204	-489.0075

Source of Variation	Σy^2	Σxy	Σx^2	$\Sigma(xy)^2/\Sigma x^2$	Adjusted Values d.f.	Adjusted Values Sum of Squares	Adjusted Values Mean Square	F	Hypothesis Tested*
Between Groups	64.0896	76.4128	126.9827		2	114.0728	57.0364	4.492	Reject $.05 > P > .01$
Within Groups	1,150.1492	-489.0075	1,625.1901	147.1387	79	1,003.0105	12.6963		
Total	1,214.2389	-412.5946	1,752.1830	97.1556	81	1,117.0833			

*Null Hypothesis

Code: 1 = 1 year; 2 = 2-5 years; 3 = 6-10 years; 4 = 11-19 years; 5 = 20 or more years case. The relationship between class gains and the teacher's number of years of teaching experience was found to be significant at the .05 level (analysis of variance and covariance). Information on this analysis will be found in Table I.

The fact remains that most of the pupils did gain significantly on test scores in the five-month interval between the tests, but some fifth-grade groups gained more than others, some did not gain significantly, and some lost rather than gained. These differences in amounts of gain are not therefore due to the lapse of time and increasing maturity of the pupils alone; neither do they appear to be significantly related to pupil I. Q. or to any of the teacher or teaching-situation factors which were investigated, with the exception of the teacher's years of teaching experience.

Further analysis could be made to determine whether any combination of teacher-factors showed a significant relationship to pupil gains, or whether reading test results showed a significant relationship to science test results. These analyses are not reported here.

The results of the statistical analysis of questionnaire and test data may indicate a true lack of relationship between the teacher-factors checked and the pupils' science achievement, or they may reflect the fact that the general level of teacher preparation in science was limited; therefore, the differences between teachers and teaching practices were not marked enough to effect significant differences in class achievement.

The study reveals some information which has implications for the preparation of elementary school teachers.

In the group of 1850 fifth-grade pupils on whom complete test data were available, 147 were found to have I. Q.'s of over 120. It can thus be estimated that approximately 1732 Minnesota fifth-grade pupils (1954), exclusive of those in Minneapolis, St. Paul, and Duluth, had I. Q.'s of 121 plus. With the current need for capable adult scientists, and the knowledge that they are generally from this high I. Q. group, the indications from this study that many elementary school teachers are failing to (a) differentiate classroom work sufficiently so that the talented achieve in proportion to their abilities, and (b) sample adequately the various science areas and activities so that vocational and avocational interests are aroused early, have important implications. Certainly the tendency to emphasize biological science at the expense of physical science, the limited physical science background of the typical teacher in this survey, the predominant use of the textbook-discussion method in teaching science, and the almost complete absence of individual laboratory work and true experiments in the classrooms, indicate that the potential scientists are not being provided for.

The objectives of elementary school science, moreover, do not apply only to those who intend to pursue scientific careers; they apply to all children. It is unlikely that Minnesota children in general are profiting any more from the present science program than are the high-ability children. Research is needed to determine the best ways of providing prospective and in-service teachers with the "know-how" and emotional push necessary to enable them to provide a dynamic science program for each one of their pupils, potential nuclear physicist and bird-watcher alike.

This study made possible a detailed description of the teachers, science teaching practices, and science teaching facilities in Minnesota fifth-grade classrooms in 1954. It made available a record of the I. Q.'s and science test scores of a representative sampling of Minnesota fifth-grade classes, from which it was possible to generalize for the state as a whole. It provided information on the extent to which the objectives of elementary school science, as measured by gains in a science achievement test, are being achieved in the elementary schools of Minnesota, the limitations in the present program, and suggestions for further research on the best methods of preparing teachers to further the achievement of the objectives of elementary school science with their pupils.

FOOTNOTES

1. Jane Johnston. The Relative Achievement of the Objectives of Elementary School Science in a Representative Sampling of Minnesota Schools, unpublished Ph. D. thesis, University of Minnesota, 1956.

2. A Guide for Instruction in Science and Conservation, Curriculum Bulletin Number 7 (St. Paul, Minnesota: Minnesota State Department of Education, 1951).

3. Palmer O. Johnson. Statistical Methods in Research (New York: Prentice Hall, Inc., 1949), pp. 126-127.

4. Ibid., p. 78.

5. Palmer O. Johnson. Cluster Sampling, unpublished material, University of Minnesota.

CONTROLLED EXPERIMENTATION IN THE CLASSROOM*

JULIAN C. STANLEY
University of Wisconsin

MY THESIS is that much more controlled experimentation under classroom conditions should be done. "Experimentation" as used here is not meant to include uncontrolled experientation, important though such experiental analysis undoubtedly is. Nor do I include status studies unless they are imbedded in an experimentally controlled design.

There are two general types of controlled experimentation in classrooms. One of these is the "methods" study in which two or more ways of doing something are compared in an unbiased fashion. The other is so-called "fundamental research," whose intent is to derive general principles applicable beyond the immediate situation in which they are found. We need far more of both types of research.

The Speech Study

Consider the following example of the many methodological decisions teachers must make virtually every day. Recently I had the pleasure of helping several high-school teachers set up an experiment to determine the effectiveness of observing seventh graders unobtrusively in the natural context of the classroom and then telling them in writing specifically which speech faults they needed to correct, and how to correct them. Logically and intuitively, it was apparent to the speech teachers that this procedure should be quite effective, far better than just instructing students in the speech class or having no formal instruction at all. They were scientifically minded, however, and therefore decided to subject their hunch to empirical test. The design of their study should serve to illuminate several aspects of careful experimentation.

They might have taken the students in each speech class at the beginning of the semester and divided them randomly into two groups, one of whom would be "followed" in other classes during the semester and notified of speech imperfections. The other group would not be "followed." If the extra attention was effective in improving speech behavior, then the "followed" group should be the better speakers and oral readers by the end of the semester. Having both groups in intermixed order speak and read orally for judges who do not know to which group a given individual belongs is a means of securing unbiased ratings that can then be separated for the two groups and compared. If the average rating of the "followed" group is significantly larger than the average rating of the "not followed" group, then we would conclude that the extra attention is having some positive influence upon speaking. Of course, the influence still might not be great enough to warrant the expenditure of time and money, but at least "following" would be shown to be having some effect.

In this sort of experiment it is easy to test additional hypotheses without using any more students than are necessary for the simple comparison of the "followed" and "not-followed" group. One might, for instance, hypothesize that "following" would work better for girls than for boys. By dividing the students first by sex and then randomly by followed vs. non-followed group (see Table I), we can answer two new questions: (1) Do boys differ from girls on the ratings? (2) Are there differences between the sexes with respect to the value of "following"?

Going still further, we might include in the experiment both those persons who are taking speech and those taking English, dividing all the speech students into two sex groups and subdividing each of these into the "followed" vs. the "non-followed." Similarly, we would separate the English students into boys and girls, then dividing the boys into "followed" vs. "non-followed" and the girls likewise. See Table II.

This arrangement results in eight groups: boys taking speech who are followed, boys taking speech who are not followed, girls taking speech who are followed, girls taking speech who are not followed, boys taking English who are followed, boys taking English who are not followed, girls taking English who are followed, and girls taking English who are not followed. It enables us to answer seven different questions, contrasted with only one for the two-group design and three for the four sex-method groups, and yet requires no more students than the two-group setup. The seven questions that can be answered are: Do boys differ from girls with respect to speaking or oral reading ability? Do those persons taking speech differ from those taking English? Does the "followed" group differ from the "non-followed" group? Are there interactions of sex, subject, and method, sex and subject, sex and method, or subject and method?[1]*

---

*Revised version of a paper read to the Educational Research Section of the Wisconsin Education Association, November 4, 1954.

The four- and eight-group arrangements shown in Tables I and II are both "factorial designs." We have two levels of each factor—sex, subject, and method—combined in all possible ways with each other to yield eight different sex-subject-method groups in Table II. The huge advantages of factorial designs are that they allow several variables to be manipulated simultaneously and then to be evaluated independently of each other and that they make the testing of interactions possible. Frequently heard objections to the artificiality of controlled experimentation involving the manipulation of a single variable in a sort of vacuum, with other factors held as constant as possible, are not applicable to the factorial design. It permits us to approximate a natural setting in our experimentation, while at the same time testing several possible factors with nearly as much efficiency as if we were considering only a single variable such as "followed" vs. "non-followed." As yet, this design is seldom employed in educational research, though it is of wide utility.

In Table II we have the outline of a methods experiment actually conducted. Results were all negative. The boys and girls spoke and read equally well, the speech classes did not do appreciably better than the English classes, the "followed" students were no better speakers or readers at the end of the semester than the "non-followed" group, and none of the four interactions was significant. Therefore, within the limitations of our experimental procedure we are forced to conclude at least tentatively that "following" students is a waste of time and money.

It is essential to remember the conditions under which we experimented, however. Having used only seventh graders in a certain rather atypical type of school, we cannot generalize our conclusions to other grades or other schools. The method of "following" we used is the only one for which we have information; another method might be effective. Our findings are limited by the nature of the speech and oral reading ratings secured, and by the competencies and idiosyncrasies of the 10 particular raters involved. Furthermore, the rating situation itself may be too formal and unlifelike to elicit natural speech behavior from the students. Finally, we gathered our data at the end of the first semester and consequently cannot make statements applicable to the whole year. Nevertheless, there is certainly no positive evidence in this carefully controlled study that being followed is worthwhile. Henceforth, the burden of proof is upon proponents of "following" rather than upon its opponents. They may want to launch new experimentation to test further hypotheses, such as the following: another well defined method of following is effective; following works better in other grades than it does in the seventh; following is more effective in other schools than here; other measures of speaking and oral reading ability reveal the effectiveness of following.

Where possible, we try to design methods experiments so as to answer questions more general than the purely local problem under consideration. Relating the study to a conceptual scheme or theory from which results can be deduced in advance enables us to test important aspects of the theory while we are getting our specific information. How well this can be done depends largely upon how advanced conceptually the field in which we are working is. If speech theorists have devised definite frameworks to unify their field, then we can test deductions based upon these. For example, if one theory emphasizes changing the individual's self-concept in a certain way, we can use a method of "following" that is congruent with it. Or, if another theorist emphasizes giving the student successful experiences in a speaking situation, we can try to accomplish that. Because of the numerous ways to assist or tutor a student, we must select the ones for trial on the best possible logical and empirical grounds, rather than just picking at random.

The Opinion-Change Study

The methods experiment of Table II constituted action research, since it was designed by actual teachers of speech to determine whether or not they should continue following their students. Its results were of immediate practical value to these teachers in their school. Some research is less related to action and more to general speculation. A study conducted by Herbert Klausmeier and myself[2] may illustrate this semi-theoretical type of experiment.

We were interested in whether or not role playing changes extreme opinions. Our subjects were 145 beginning graduate students, 62 in my advanced educational psychology class and 83 in his. Thirteen days before the beginning of the experiment itself we obtained ratings of attitude toward world government on a four-question opinionnaire. Then for my experiment I chose low scorers ("isolationists") and high scorers ("internationalists"), leaving the middle scorers to rate secretly the quality of the role playing. One-fourth of the extreme scorers were assigned to play a role opposite to their initial opinions,

---

*All footnotes will be found at end of article.

TABLE I

A FOUR-GROUP (SEX-METHOD) FACTORIAL DESIGN*

Sex	Method	
	Not Followed	Followed
Boys		
Girls		

*There are n students in each of the four groups, or $4n$ students in the entire experiment. Half of the $2n$ boys are assigned *randomly* to the “not followed” group and the other half to the “followed” group. Similarly, half of the $2n$ girls are assigned *randomly* to the “not followed” group.

TABLE II

AN EIGHT-GROUP (SEX-SUBJECT-METHOD) FACTORIAL DESIGN*

Sex	Subject	Method	
		Not Followed	Followed
Boys	Speech		
	English		
Girls	Speech		
	English		

*Each boy is assigned *randomly* to one of four groups (speech-not followed, speech-followed, English-not followed, or English-followed). Likewise, each girl is assigned *randomly* to one of the four groups. Preferably, each of the eight groups will contain *the same* number of students; in the speech study there were six students in each of the eight groups.

one-fourth play a role congruent with their initial opinions, one-fourth to play no role but to serve as first-row observers, and one-fourth to go to the library before the experiment began and work on unrelated material. Thus there were four "role" groups: different role, same role, observer, and outside control. Each of the four role groups contained an equal number of men and women, and each had half of its persons whose initial opinions were "isolationist" and half who were initially "internationalist." In all, there were 16 role-sex-initial opinion groups, as shown in Table III. Seven hypotheses can be tested with this design, using change-of-opinion scores based upon differences between the first opinionnaire and a similar opinionnaire administered the day after the experiment:

Are the differences among the four roles significant? Do the changes of men differ from those of women? Does it make any difference with respect to opinion change whether the person was originally "isolationist" or "internationalist"? Are any of the interactions among sex, initial opinion, and role significant?

Klausmeier's design was identical with mine, except that where I used a "sex" classification he used "age": 31 or less vs. 32 or more. Under our conditions, role playing had no effect upon opinion. We had hypothesized that persons playing a role in line with their initial opinions should change little, while persons playing an opposite role should shift in the direction of that point of view, but this did not happen. Nor did the persons playing a compatible role change significantly less than the persons who left the room without even observing the role playing. Our only positive finding was a slight tendency for the women in my study to get more world-minded and the men less world-minded; this interaction of sex with initial opinions may be attributable merely to chance.

Implications

Since both of the studies I have outlined yielded essentially negative findings, you may wonder whether these experiments were worth performing. Obviously, it is valuable to know that following students to observe and record their speech errors for their information is not effective. It is probably of even greater importance to learn that confident claims for the opinion-modifying influence of role playing do not hold up in a controlled experiment.[3] Both experiments highlight the fact that hypotheses must be tested empirically as well as logically. Our best brain-children may be worthless or downright pernicious. Expert opinions, pooled judgments, brilliant intuitions, and shrewd hunches are frequently misleading. Ultimately, they must be tested by the careful gathering of evaluative data if education is to advance on the basis of sound principles.

We badly need educational experimentation of the controlled variety. Too much of our research effort has been directed toward questionnaires, statistical compilations, and dead-end correlational studies. We have shied away from designing and executing studies involving the manipulation of pertinent variables. Our colleagues in agriculture and the behaviorial sciences such as psychology have forged far ahead of us experimentally. Yet we cannot rely upon psychologists, sociologists, and anthropologists to do our experimentation. They provide much relevant theory and many enticing hypotheses for us, but we must hoe our own garden; they cannot be expected to do this.

Obstacles to Controlled Experimentation

In my opinion, one of the major reasons why experimentation in education has languished is that our graduate schools have been confused by a false practitioner versus research distinction. This has lead in most schools of education to the training of few, if any, persons in educational experimentation as such. Whereas almost no psychology major can earn a Ph. D. degree in a reputable university without prolonged exposure to experimental psychology, statistics, and measurement, we let the great majority of our doctoral candidates get through their dissertations with hardly the basic rudiments of training for experimentation. And training is essential. While it is true that many persons who receive such training will actually engage in little experimentation, untrained individuals are virtually certain to do none or to do it ineptly.

Another reason for the paucity of educational experimentation is that university professors in schools of education typically do very little themselves. Until they provide the training for experimentation and set appropriate examples for their students, improvement is unlikely.

Few public school administrators and even fewer parents realize the dire necessity for continual experimentation. Some equate experimentation with vivisection. Others protest against the use of control groups, saying that if a given method is likely to be of value they want all the students to have it, not realizing that in this way they make it difficult, if not impossible, ever to assess the worth of the method. Perhaps the widely publicized tests of Salk's polio vaccine will be a salutary antidote to this attitude, though it is doubtful that many persons not already aware of the need for controls in experimentation will perceive the similarity between the polio study and, say, the introduction of a new method for teaching handwriting.

TABLE III

A FACTORIALLY DESIGNED CHANGE-OF-OPINION EXPERIMENT INVOLVING 16 GROUPS*

Sex	Initial Opinions	Role			
		Congruent with Initial Opinions	Opposite to Initial Opinions	First-Row Observer	Outside Control
Men	Internationalist				
	Isolationist				
Women	Internationalist				
	Isolationist				

*There were two persons in each of the 16 groups, a total of 32 in the experiment, not counting the 30 raters. Each man was assigned randomly to one of the four role groups; women were assigned similarly.

Two essential ingredients of controlled experimentation are difficult to incorporate into our thinking. One is the imperative need for designing experiments "from scratch." Experiments do not "just grow." They must be planned in advance down to the finest detail. The final analysis of data must be anticipated and worked through abstractly or in a pilot study before the experiment itself begins. This preplanning phase will usually require more time than the actual experiment. Once all aspects of the design have been thought through and refined as much as possible, the experiment must be run off exactly as planned, allowing only for unavoidable disasters. We hear a great deal about "flexible experimentation," but once set in motion a controlled experiment is completely rigid and inflexible. Flexibility must be reserved for the preplanning and postplanning phases. The experimental findings almost invariably suggest new hypotheses that can be tested experimentally.

The second difficult concept to put over is the crucial need for randomization. The very notion of randomization is hard to grasp, since it implies chaos for the single case and a high degree of order for the group. Suppose that you have 30 pupils in a class and that you want a control group of 15 and an experimental group the same size. How shall you divide the class into halves in order not to give an undeserved lead to one group or the other? If there were five seats to a row, you might consider taking all students in the first three rows for one group and all in the last three rows for the other group. Would this procedure yield two groups of equal ability for, say, an arithmetic experiment? Not in most classes, for seating is hardly likely to be random. If mostly girls sit near the front of the room and mostly boys near the rear, and if one sex surpasses the other in arithmetic skill, results of the experiment will be biased at the very beginning.

We might put all 30 names into a hat, each on a separate slip, shuffle them thoroughly, and then draw out 15. Or, somewhat less defensibly, we might alphabetize the 30 names and put the odd-numbered persons (1, 3, 5, . . , 29) in one group and the even-numbered persons (2, 4, 6, . . . , 30) in the other. The best method is to number the children from 1 through 30 (or from 0 through 29) and draw 15 numbers within this range from a table of random numbers. In all three instances—shuffling slips, alphabetizing, and random numbers—it is essential to decide in advance which will be the control group and which the experimental. We cannot wait until we note into which group our prize pupil happens to fall and then call this the experimental group. Experiments must be without bias if they are to yield unambiguous conclusions.

I cannot stress too heavily the need for getting fresh data specifically designed to answer the questions you pose. Too often a researcher, particularly a thesis-driven graduate student, is tempted to exhume some moribund material he finds stored away in a convenient mausoleum and perform a statistical autopsy on it. This "seeing what the data died of" is good clinical medicine but rarely worthwhile research. The persons who originally collected the data did not have the present problem in mind, so it would be quite unusual for the data to bear directly on the current hypotheses. About all you can get out of this post-mortem analysis is a crazy patchwork quilt of guarded conclusions and tortuous qualifying statements. The customary upshot is to say that "more research is needed." Of course it is, but not research of this foredoomed variety.

Almost as culpable is the person who gathers mountains of data as he goes along, without any particular experimental design, in the hope that eventually it will serve to test some hypotheses he gets around to formulating. In many instances this is a collossal waste of time and energy, harking back to the exploratory days of the late nineteenth century and the biometrical approach of Karl Pearson. Since 1925, and particularly since 1935, when R. A. Fisher's The Design of Experiments appeared, we have known about the inefficiency of this "sawed-off shotgun" approach. No longer do we need huge numbers of experimental subjects in order to reach valid conclusions. Small-sample theory enables us to emphasize the rigor of the experimental design instead of the size of the sample itself. Ten thousand subjects in an ill-planned study may yield results more equivocal than 40 subjects in a carefully designed experiment.

Preparation for Classroom Experimentation

There are many experiments you can do in the classroom. Most of these probably fall within the category of action-method studies, but some may be of more general import. Remember that the generality of your findings is inextricably bound up with the formulation of the problem; if you strive for generality of application rather than viewing your investigation as answering a merely local problem, you can make a contribution beyond the confines of your school.

You will want to have the necessary training in statistics (particularly the analysis of variance) and research methods before trying to design an experiment, and/or you will want to work closely from the beginning with a specialist who understands these matters well. Unfortunately, there are not many such persons available, perhaps no more than 100 educationalists in the whole United States. A possible solution is to take statistics and experimental design courses in the psychology department of a major university. Collaboration is the desirable plan, however, even though you do acquire considerable technical sophistication.[4] Cooperative efforts such as the speech study I described earlier benefit all persons involved immensely.

In any event, it is essential that you design your experiment thoroughly before beginning to "turn the crank" and that all collaboration be started at the very begining, rather than after the data are in. As the statistician must nearly always say to the researcher who consults him too late, after the experiment has been mismanaged badly, "Of all sad words of tongue or pen, the saddest are these, it might have been." Ordinarily, the specialist in experimental design does not know your subject matter field as well as you, but he can point out logical flaws and methodological imperfections that might nullify all your otherwise commendable efforts.

Recapitulation

This paper has been concerned entirely with controlled experimentation in the classroom, which I believe we have neglected to our great detriment. Most decisions about methods have been based upon colloquial, anecdotal, or administrative considerations rather than experimentation. Seldom are adequate control groups incorporated into classroom experiments. The necessity for long-range experimental design is not usually appreciated by teachers and administrators. The principle of randomness is often misunderstood or ignored in favor of elaborate matching, which has several disadvantages. Worst of all, few teachers, including those with doctoral degrees, get even minimal training for modern experimentation. Our professional literature is virtually devoid of well controlled experimental studies in the classroom. We continue to pool ignorance via conferences, questionnaires, rating scales, opinionnaires, and ineffective correlational studies, all of which are valuable for certain purposes but not sufficient in themselves.

If we are to advance beyond the dark ages of educational pre-science, we must emulate the experimental proficiency and zeal of colleagues in other behaviorial sciences. Ours is a distinctive task; education has its own problems to solve. Only by close collaboration between experimentally aware persons at all levels in a widespread research partnership can we hope to make headway against the vast complexities of the teaching-learning process.

FOOTNOTES

1. If there are no significant interactions, the "main effects" are additive. Suppose, for example, that girls are superior to boys, persons taking speech are superior to those taking English, and the "followed" group is superior to the "non-followed" group. Then, if interactions are nil, the best speakers will be the girls taking speech and being followed. The worst speakers will be the boys taking English and not being followed. However, if, despite the general superiority of girls, speech, and following, the boys taking English and not being followed were the best speakers, we would have a triple interaction (sex, speech, and method).

   Two-factor interactions may also occur. Girls taking English might be superior to girls taking speech, despite the overall superiority of the speech group over the English group. Also, boys being "followed" might be better speakers than the followed girls, despite the overall speaking superiority of the girls. Or those persons taking English and being followed might be better speakers, regardless of sex, than those taking speech and being followed, even though for the experiment as a whole the speech group is superior to the English group.

2. "Opinion Constancy After Formal Role Playing," Journal of Social Psychology (in press).

3. As in the speech study, one must not overgeneralize the results of the Stanley-Klausmeier investigations. Definite limitations were set by the type and small number of subjects used, the formal nature of the role playing situation, and the opinion area involved. Janis and King, studying "The Influence of Role Playing on Opinion Change," [Journal of Abnormal and Social Psychology, XLIX (1954), pp. 211-18] with small groups under non-classroom conditions, found evidence for such change. For a recent review article that contains 22 references, see J. H. Mann, "Experimental Evaluations of Role Playing," Psychological Bulletin, LIII (1956), pp. 227-34.

4. Even the experimental-design "experts" in education and psychology consult each other and the mathematical statisticians frequently.

RECENT TECHNIQUES FOR ANALYZING ASSOCIATION IN CONTINGENCY TABLES AS APPLIED TO AN ANALYTICAL FOLLOW-UP SURVEY OF EDUCATION GRADUATES*

SAMUEL T. MAYO
Loyola University
Chicago, Illinois

Introduction

THE PURPOSE of the present paper is to describe some recent techniques for analyzing association in contingency tables and to illustrate their application to an important educational problem evaluating the human product relative to the occupational destination of graduates of a teacher-training institution.

The kinds of statistical techniques used in the present investigation are those applicable to enumerative data in the form of contingency tables, which are classified on the basis of attributes. The concept of attributes will be considered elsewhere in this manuscript. The use of contingency tables to analyze categorical data is not new, dating back some forty-odd years to the work of Karl Pearson. However, little fundamental work on the improvement of the theory and applications of contingency tables seems to have been done until just after World War II. From that time a number of significant techniques have been developed by statistical workers in the fields of biometrics and agriculture which promise to greatly extend the usefulness of the contingency table. It should be recognized that there is no reason, in principle, why such techniques should not find great usefulness in the behavioral sciences as well.

Chi-square, as usually employed, has been traditionally limited to contingency tables classified on the basis of only two attributes at a time and consisting of a "large" sample. Among the resulting disadvantages of this more familiar application of contingency tables are these:

1. The chi-square test for a contingency table suffers from the weakness that it covers all forms of departure from expectation and is correspondingly insensitive to departures of a specified type.
2. Two-way tables, analyzed piecemeal, ignore higher-order interactions, which involve three or more attributes simultaneously.
3. "Large" samples are not always available in behavioral research.

Some recently developed techniques considered here serve to overcome these weaknesses. Among these techniques are:

1. Approximate tests of significance for contingency tables in which association is known to exist, so that the significance of departures of a specified type (e.g., correlation or regression) may be tested;

2. Approximate tests of significance for complex contingency tables involving various hypotheses among as many as four attributes simultaneously (i.e., tests of the significance of higher-order interactions);

3. Exact tests of significance for "small" sample data, or contingency tables with "small" or zero theoretical cell frequencies and classified on the basis of two or three attributes.

The first of these types of techniques, those by which one may investigate departures of a particular type, utilize scoring schemes for rows and columns of the contingency tables. Such techniques have been described by Cochran (2), Stuart (11), Yates (13), and Williams (12). The techniques of Yates and of Williams will be described and illustrated in the present paper.

In the second type of technique, or those for complex tables for which one may test hypotheses involving three or more attributes simultaneously, chi-square may be estimated by either of two approximate solutions. In one solution as given by Bartlett (1), Norton (9), and Snedecor (10), chi-square as a test for higher-order interaction in a $2 \times 2 \times 2$ table or the general 2^n form can be approximated by solving a higher-order equation. Thus, in the $2 \times 2 \times 2$ case, one would solve a cubic equation. In the other type of solution, chi-square for an $r \times s \times t \times u$ (or four-way) contingency table with any number of categories per attribute can be estimated by application of the likelihood ratio criterion as described by Mood (8) and Doi (3).

In the third type of technique, or those comprising exact tests, one considers the exact relative frequency of a selected configuration of cell frequencies among all configurations deduced from the observed marginal totals, which are used as ancillary statistics. The concept was first introduced by Fisher (4) for the 2 × 2 case and has been extended to more complex cases by Freeman and Halton (5).

---

*The author wishes to acknowledge the help of his advisor, Dr. Palmer O. Johnson, in the carrying out of the investigation upon which this paper is based.

These techniques will now be described in detail and illustrated elsewhere in the manuscript by the survey data.

Scoring Techniques in Simple Contingency Tables

The significance of association between two attributes in a simple (or two-way) contingency table can be tested by the usual chi-square test. However, chi-square is an overall test of all forms of departures from expectation. Under the null hypothesis that the two attributes are not associated, the expected frequencies in any row (or column) are proportional to the marginal frequencies. When one rejects the null hypothesis of independence and association is assumed to exist between the attributes, it is of practical importance to be able to specify the nature of the association. Williams (12) maintains that the interpretation of the data is simplest if the association can be accounted for in terms of the correlation between a single pair of variates corresponding to the two attributes. In order to make this interpretation, one must make the assumption that both attributes are based on characters which are either directly quantitative or are in the form of gradings which can be regarded as having an underlying quantitative basis. Williams shows that significance tests, developed for discriminant analysis and for the interpretation of interactions, exact when the variates involved are normally distributed, may be applied as tests asymptotically exact to contingency tables.

The methods of Yates (13) and Williams (12) differ in the manner of determining the scores for rows and columns. In the Yates' method the scores are chosen arbitrarily. Yates has pointed out that any system of scoring may be assigned to each of the classifications. In the Williams' method, on the other hand, scores are derived empirically from the data and must satisfy certain identities. The empirically-derived scores serve to maximize the correlation coefficient obtainable from a given set of data. Thus, they can be brought into closer concordance with the data.

The Yates' Method of Scoring Simple Contingency Tables—The symbolic notation of Yates was modified in order to be more consistent with that of Williams as well as with the usual statistical symbols. The notations used here were as follows:

n_{ij}	observed frequency in ith row and jth column
$n_{i.}$	total frequency in the ith row
$n_{.j}$	total frequency in the jth column
x_i	arbitrary score assigned to the ith row
y_j	arbitrary score assigned to the jth column
B	numerator of the formula for regression coefficients of y_j on x_i and x_i on y_j
A	denominator of regression formula for y_j on x_i
A'	denominator of regression formula for x_i on y_j
b_{yx}	regression coefficient of y_j on x_i
b_{xy}	regression coefficient of x_i on y_j
r	number of rows in contingency table
s	number of columns in contingency table
α	significance level for a particular t

The Yates' symbolism for a 3 × 3 contingency table with assumed arbitrary scores being the series -1, 0, and +1 for both rows and for columns is given in Figure 1.

The quantities A, A', and B are calculated by the following expressions:

$$A = n_{..} \sum_i x_i^2 n_{i.} - (\sum_i x_i n_{i.})^2$$

$$A' = n_{..} \sum_j y_j^2 n_{.j} - (\sum_j y_j n_{.j})^2$$

$$B = n_{..} \sum_i (x_i \sum_j y_j n_{ij}) - (\sum_i x_i n_{i.})(\sum_j y_j n_{.j})$$

The two regression coefficients are given by

$$b_{xy} = B/A$$

and

$$b_{yx} = B/A'$$

To test whether each regression coefficient is significantly different from zero, we first compute the standard error of each regression coefficient and then apply the t test. The standard errors of the coefficients are given by

$$\text{S.E. of } b_{xy} = \sqrt{A'/An_{..}}$$

and

$$\text{S.E. of } b_{yx} = \sqrt{A/A'n_{..}}$$

and the corresponding t ratios are given by

$$t = \frac{b_{xy}}{\text{S.E. of } b_{xy}}$$

and

$$t = \frac{b_{yx}}{\text{S.E. of } b_{yx}}$$

The t's are not differentiated by subscripts, since they will be numerically equal when applied to a given set of data. The obtained t's are referred to either the t table with infinite degrees of freedom or to the normal probability tables.

After we have estimated the magnitude of the effect of one variate upon the other in the parent population, we may further inquire as to the errors of this estimate. Fiducial limits for the total variation of y_j accounted for by x_i between extremes of both variates may be estimated by multiplying b and its standard error by $(r - 1) = (s - 1)$ first, then computing

$$b \pm t_{\alpha} \text{ (S. E. of b)}$$

An alternative test of significance of the association of row and column scores is given as a chi-square with 1 degree of freedom by

$$X^2 = B^2 n_{..}/AA'$$

However, if the t's for testing the significance of regression coefficients have already been calculated, we can obtain this same value of chi-square by merely squaring t.

The Williams' Method of Scoring Simple Contingency Tables—The symbolic notations for the Williams' method are the same as for the Yates' method, with the following exceptions:

x_i	denotes empirically-derived score for the ith row
y_j	denotes empirically-derived score for the jth column
r_{xy}	denotes the correlation between the row scores and column scores

The matrix notations will be explained in the text as they appear. Further explanation of matrix symbolism and algebraic rules may be found in texts by Gulliksen (6:160), Mood (8:170-176) and by other authors.

As in the Yates' method, the first step in the Williams' method is to make an overall test of association by the usual X^2 test. This is done in order to insure that association does, in fact, exist, since either the Yates' or Williams' method is primarily an aid to interpretation of a contingency table where association is known to exist.

Solution of the empirical scores proceeds by a number of operations with matrices.

The original r × s contingency table with elements n_{ij} is transformed to the matrix G with elements given by

$$g_{ij} = \frac{n_{ij}}{\sqrt{n_{i.}\, n_{.j}}}$$

This amounts to dividing each cell frequency in the contingency table by the geometric mean of the row total and column total for that particular cell.

The matrix G is transformed to matrix T by the matrix multiplication

$$T = GG'$$

This formula says that matrix T is obtained by multiplying G by its transpose. The transpose of a matrix is obtained by making its rows into columns and its columns into rows. Each element in the T matrix is obtained by following the rules for multiplying two matrices. Mood (8) has stated the definition for matrix multiplication in general as follows:

> The element in the ith row and jth column of the product matrix is obtained by multiplying the elements of the ith row of the left-hand matrix by the corresponding elements of the jth column of the right-hand matrix and adding the results.

In order to calculate the scores for rows and columns of the contingency table which would maximize the correlation between the two scores, it is necessary to solve for the second largest latent root of the matrix T. We shall designate this root by the symbol λ. This root is equal to the square of the maximized correlation coefficient, and the latent vector of this root gives the required set of scores for the rows. Column scores are calculated from row scores by an appropriate formula, which will be given subsequently.

To solve for λ we obtain the characteristic equation of matrix T by first subtracting the unknown quantity λ from each of the three terms in the main diagonal of T, then by treating the resulting matrix as a determinant, and finally by expanding the determinant to yield the characteristic equation.

When the numerical value of λ has been determined to the desired degree of accuracy, the row scores are found for a set of solutions b_i to the set of simultaneous equations

$$[T - \lambda I]\left\{ b_i \right\} = 0$$

and finding a set of scores

$$x_i = kb_i$$

proportional to the b_i's so that the restrictions

$$\sum_i n_{i.} x_i = 0$$

and $$\sum_i n_{i.} x_i^2 = n_{..}$$

are satisfied.

The column scores are calculated by the relation

$$y_j = \sum_i n_{ij}x_i/r_{xy}n_{.j}$$

and the obtained scores must satisfy the restrictions

$$\sum_j n_{.j}y_j = 0$$

and

$$\sum_j n_{.j}y_j^2 = n_{..}$$

A check on the accuracy of the scores is given by the formula for the correlation in terms of row and column scores as given by

$$r_{xy} = \sum_j \sum_i n_{ij}x_iy_j/n_{..}$$

which should be numerically equivalent to $\sqrt{\lambda}$.

Williams showed that the formula for testing the significance of the association of row and column scores which was given by Yates as

$$X^2 = B^2n_{..}/AA'$$

is equivalent to testing

$$X^2 = n_{..}\,r_{xy}^2$$

as a chi-square with one degree of freedom. Since r_{xy} is also a correlation coefficient, an alternative test is to treat

$$F = \frac{(n_{..} - 2)r_{xy}^2}{1 - r_{xy}^2}$$

as an F with 1 and $n_{..}$ -2 degrees of freedom.

Higher-Order Interactions in Complex Contingency Tables

Whereas in a two-way contingency table, we test the hypothesis that there is no association between the two attributes of classification, which is equivalent to saying that there is no first-order interaction between the attributes—in a complex table, we may test the hypothesis that there is no higher-order interaction among the several attributes. Several different kinds of higher-order interactions may exist, and later these will be considered when the likelihood ratio criterion is applied. First, however, the 2 × 2 × 2 case will be considered.

Second-Order Interaction in the 2 × 2 × 2 Contingency Table—The method of testing second-order interaction for the 2 × 2 × 2 contingency table is due to a method suggested by Bartlett (1), developed by Norton (9), and illustrated by Snedecor (10). The symbolism of these authors is followed in the present paper.

Consider, first, the two fourfold tables with cell contents a, b, c, d and a', d', c', d', respectively, as shown in Figure 2, where the notations are given as follows:

A, B, C	denote the three dichotomized attributes
i, j, k	denote numbers of the categories of the three attributes, respectively, which take on values of 1 or 2
a, b, c, d	denote the cell contents of the first 2×2 table of attribute C
a', b', c', d'	denote the cell contents of the second 2 × 2 table of attribute C
p_a, p_b, p_c, p_d	denote the parameter values of the probabilities of a case from sample N falling in the indicated cell
$p_{a'}, p_{b'}, p_{c'}, p_{d'}$	denote the parameter values of the probabilities of a case from sample N' falling in the indicated cell

Each of the 2 × 2 tables which is classified by the two attributes A and B may be examined for independence separately by the usual chi-square test. In these two tables the null hypotheses tested are

$$p_ap_d = p_bp_c$$

and

$$p_{a'}p_{d'} = p_{b'}p_{c'}$$

respectively.

In the theory being followed, the sample estimate of first-order interaction is the ratio

$$a/b : c/d = a/c : b/d = ad/bc$$

Aside from the continuity adjustment and from errors of sampling, if the null hypothesis is true ($X^2 = 0$), then $ad/bc = 1$. In this sense the hypothesis tested by chi-square is that in the population the interaction ratio is unity.

In the 2×2×2 table, the interaction in one 2×2 table may be different from that in the other. The ratio of these first-order interactions is known as the second-order interaction. The sample estimate of the second-order interaction is

$$\frac{ad/bc}{a'd'/b'c'} = adb'c'/a'd'bc$$

FIGURE 1

SYMBOLISM FOR YATES' METHOD WITH AN EXAMPLE OF NUMERICAL SCORES

		(Score)^2	1	0	1			
		Num. Score	-1	0	+1			
		Sym. Score	y_1	y_2	y_3		Total Score	Mean Score
(Score)^2	Num.	Sym.						
1	+1	x_1	n_{11}	n_{12}	n_{13}	$n_{1.}$	$\sum_j y_j n_{1j}$	$\sum_j y_j n_{1j}/n_{1.}$
0	0	x_2	n_{21}	n_{22}	n_{23}	$n_{2.}$	$\sum_j y_j n_{2j}$	$\sum_j y_j n_{2j}/n_{2.}$
1	-1	x_3	n_{31}	n_{32}	n_{33}	$n_{3.}$	$\sum_j y_j n_{3j}$	$\sum_j y_j n_{3j}/n_{3.}$
			$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{..}$		
		Total Score	$\sum_i x_i n_{i1}$	$\sum_i x_i n_{i2}$	$\sum_i x_i n_{i3}$			
		Mean Score	$\frac{\sum_i x_i n_{i1}}{n_{.1}}$	$\frac{\sum_i x_i n_{i2}}{n_{.2}}$	$\frac{\sum_i x_i n_{i3}}{n_{.3}}$			

FIGURE 2

SYMBOLISM FOR A 2 × 2 × 2 CONTINGENCY TABLE

		C					
	k =	1			2		
		B			B		
	j =	1	2		1	2	
A	i = 1	a	c	m	a'	c'	m'
	2	b	d	n	b'	d'	n'
		r	s	N	r'	s'	N'

Before proceeding to the presentation of the test of significance for second-order interaction, it will be well to explain the concept of deviations in 2 × 2 tables. In a 2 × 2 table the deviation (or difference between observed and theoretical frequency) is the same in absolute magnitude for all four cells. Let this deviation be symbolized by

$$x = f_o - f_t$$

where f_o denotes the observed cell frequency and f_t the theoretical or expected cell frequency. The algebraic sign of x is positive for one pair of diagonal cells and negative for the other pair, so that

$$\sum_{j}^{2} \sum_{i}^{2} x_{ij} = 0$$

It is easy to construct a table of expected numbers in which the interaction ratio is 1, as follows:

$$\begin{matrix} a - x & b + x \\ c + x & d - x \end{matrix}$$

The quantity x for a given 2 × 2 table could be computed by solving the equation

$$\frac{(a - x)(d - x)}{(b + x)(c + x)} = 1$$

which may also be expressed in the form

$$(a - x)(d - x) = (b + x)(c + x)$$

By analogy it may be shown that there is a deviation x common to all eight cells of a pair of 2 × 2 tables considered jointly as a 2 × 2 × 2. The analogous equation to solve is the cubic

$$(a - x)(d - x)(c' - x)(b' - x) = (c + x)(b + x)(a' + x)(d' + x)$$

This equation is easily remembered by a rule of thumb which indicates that in the left-hand member the letters follow a diagonal downward in the first 2 × 2 table, then across and upward in the second. Methods for approximating the real root of this cubic equation may be found in any college algebra book. Essentially, they consist of trying various values of x until two are found between which the left-hand member of the equation changes sign.

When x has been calculated to a sufficient number of decimal places, the adjusted chi-square with one degree of freedom for second-order interaction is given by the formula:

$$X^2 = (|x| - 0.5)^2 \left(\frac{1}{a-x} + \frac{1}{d-x} + \frac{1}{c'-x} + \frac{1}{b'-x} + \frac{1}{c+x} + \frac{1}{b+x} + \frac{1}{a'-x} + \frac{1}{d'-x} \right)$$

If the chi-square for second-order interaction is not significant, the pair of 2 × 2 tables may be combined into a single 2 × 2 table and tested for first-order interaction.

Likelihood Ratio Criterion for a Four-Way Contingency Table—In the foregoing section the 2 × 2 × 2, a special case of the three-way contingency table, was considered. A general means for testing several kinds of hypotheses for a four-way table by the likelihood ratio criterion will now be considered. It should be noted that the method is also applicable to three-way tables.

Consider a four-way contingency table classified by the attributes A, B, C, and D as shown symbolically in Figure 3, where the notations are given as follows:

A, B, C, D denote the four attributes
Categories of attributes are denoted by subscripts as follows:

A--(i = 1, 2, . . . , r)
B--(j = 1, 2, . . . , s)
C--(k = 1, 2, . . . , t)
D--(h = 1, 2, . . . , u)

n_{ijkh}	denotes the frequency in the cell common to the ith row, jth column, kth category of attribute C and hth category of attribute D

Marginal totals were denoted by replacing the summed index by a dot, e.g.,

$$n_{ijk.} = \sum_{h}^{u} n_{ijkh}$$

$$n_{ij..} = \sum_{h}^{u} \sum_{k}^{t} n_{ijkh}$$

$$n_{....} = \sum_{h}^{u} \sum_{k}^{t} \sum_{j}^{s} \sum_{i}^{r} n_{ijkh}$$

Probabilities for particular category values are expressed by appropriate subscripts, e.g.,

p_{ijkh}	denotes the probability of an individual's being in the ijkhth cell
p_i, p_j, p_k, p_h	denote the probabilities of an individual's being in the categories i, j, k, and h, respectively
λ	denotes the likelihood ratio

Six kinds of null hypotheses can be tested in

FIGURE 3

SYMBOLISM FOR A FOUR-WAY CONTINGENCY TABLE

			C								
			$k = 1$			2					
			B			B					
			$j = 1$	2		1	2				
$h = 1$	A	$i = 1$	n_{1111}	n_{1211}	$n_{1.11}$	n_{1121}	n_{1221}	$n_{1.21}$	$n_{1..1}$	$n_{11.1}$	$n_{12.1}$
		2	n_{2111}	n_{2211}	$n_{2.11}$	n_{2121}	n_{2221}	$n_{2.21}$	$n_{2..1}$	$n_{21.1}$	$n_{22.1}$
		3	n_{3111}	n_{3211}	$n_{3.11}$	n_{3121}	n_{3221}	$n_{3.21}$	$n_{3..1}$	$n_{31.1}$	$n_{32.1}$
			$n_{.111}$	$n_{.211}$	$n_{..11}$	$n_{.121}$	$n_{.221}$	$n_{..21}$	$n_{...1}$	$n_{.1.1}$	$n_{.2.1}$
D											
$h = 2$	A	$i = 1$	n_{1112}	n_{1212}	$n_{1.12}$	n_{1122}	n_{1222}	$n_{1.22}$	$n_{1..2}$	$n_{11.2}$	$n_{12.2}$
		2	n_{2112}	n_{2212}	$n_{2.12}$	n_{2122}	n_{2222}	$n_{2.22}$	$n_{2..2}$	$n_{21.2}$	$n_{22.2}$
		3	n_{3112}	n_{3212}	$n_{3.12}$	n_{3122}	n_{3222}	$n_{3.22}$	$n_{3..2}$	$n_{31.2}$	$n_{32.2}$
			$n_{.112}$	$n_{.212}$	$n_{..12}$	$n_{.122}$	$n_{.222}$	$n_{..22}$	$n_{...2}$	$n_{.1.2}$	$n_{.2.2}$
					$n_{..1.}$			$n_{..2.}$		$n_{.1..}$	$n_{.2..}$
			$n_{.11.}$	$n_{.21.}$		$n_{.12.}$	$n_{.22.}$				
			$n_{111.}$	$n_{121.}$		$n_{112.}$	$n_{122.}$		$n_{1...}$		
			$n_{211.}$	$n_{221.}$		$n_{212.}$	$n_{222.}$		$n_{2...}$		
			$n_{311.}$	$n_{321.}$		$n_{312.}$	$n_{322.}$		$n_{3...}$		
				$n_{11..}$		$n_{12..}$					
				$n_{21..}$		$n_{22..}$					
				$n_{31..}$		$n_{32..}$					
			$n_{1.1.}$				$n_{1.2.}$				
			$n_{2.1.}$				$n_{2.2.}$				
			$n_{3.1.}$				$n_{3.2.}$				

the four-way contingency table by the likelihood ratio. They may be stated verbally as follows:

1. The four attributes are mutually independent.
2. Three of the attributes are mutually independent.
3. Two of the attributes are independent.
4. One attribute is independent of a classification formed by combining the other three attributes into a unit.
5. One attribute is independent of a classification formed by combining two of the other attributes into a unit.
6. The combination of two attributes is independent of a combination of the other two.

Each of the six kinds of null hypotheses will be considered separately. In addition, each will be illustrated by examples of the hypotheses stated in terms of probabilities and an example of the calculation of the chi-square to test the hypotheses by an approximate method. When stating one of the hypotheses in terms of probabilities, one assumes probabilities of a case being in one category or in another category as being independent of each other. Then, by a well-known theorem of probability for compound events as shown by Mood (8:34), the null hypothesis may be stated by a proposition denoting that the compound probability (i. e., the probability of a case's being jointly in a certain pattern of categories) is equal to the product of the probabilities for the separate categories.

As an intermediate step in estimating chi-square for the various hypotheses, we calculate the likelihood ratio λ in the present method. It can be shown for complex contingency tables that in large samples the quantity $-2 \log_e \lambda$ approaches the chi-square distribution with appropriate degrees of freedom. All of the terms necessary to calculate λ for any hypothesis are given by eighteen expressions as follows:

$$\log_e \prod_{i} n_{i...}^{n_{i...}} = 2.30258 \sum_{i} n_{i...} \log_{10} n_{i...} \tag{1}$$

$$\log_e \prod_{j} n_{.j..}^{n_{.j..}} = 2.30258 \sum_{j} n_{.j..} \log_{10} n_{.j..} \tag{2}$$

$$\log_e \prod_{k} n_{..k.}^{n_{..k.}} = 2.30258 \sum_{k} n_{..k.} \log_{10} n_{..k.} \tag{3}$$

$$\log_e \prod_{h} n_{...h}^{n_{...h}} = 2.30258 \sum_{h} n_{...h} \log_{10} n_{...h} \tag{4}$$

$$\log_e \prod_{ijkh} n_{ijkh}^{n_{ijkh}} = 2.30258 \sum_{ijkh} n_{ijkh} \log_{10} n_{ijkh} \tag{5}$$

$$\log_e \prod_{jkh} n_{.jkh}^{n_{.jkh}} = 2.30258 \sum_{jkh} n_{.jkh} \log_{10} n_{.jkh} \tag{6}$$

$$\log_e \prod_{ikh} n_{i.kh}^{n_{i.kh}} = 2.30258 \sum_{ikh} n_{i.kh} \log_{10} n_{i.kh} \tag{7}$$

$$\log_e \prod_{ijh} n_{ij.h}^{n_{ij.h}} = 2.30258 \sum_{ijh} n_{ij.h} \log_{10} n_{i.jh} \tag{8}$$

$$\log_e \prod_{ijk} n_{ijk.}^{n_{ijk.}} = 2.30258 \sum_{ijk} n_{ijk.} \log_{10} n_{ijk.} \tag{9}$$

$$\log_e \prod_{ij} n_{ij..}^{n_{ij..}} = 2.30258 \sum_{ij} n_{ij..} \log_{10} n_{ij..} \tag{10}$$

$$\log_e \prod_{ik} n_{i.k.}^{n_{i.k.}} = 2.30258 \sum_{ik} n_{i.k.} \log_{10} n_{i.k.} \tag{11}$$

$$\log_e \prod_{ih} n_{i..h}^{n_{i..h}} = 2.30258 \sum_{ih} n_{i..h} \log_{10} n_{i..h} \tag{12}$$

$$\log_e \prod_{jk} n_{.jk.}^{n_{.jk.}} = 2.30258 \sum_{jk} n_{.jk.} \log_{10} n_{.jk.} \tag{13}$$

$$(14) \qquad \log_e \prod_{jh} n_{.j.h}^{n_{.j.h}} = 2.30258 \sum_{jh} n_{.j.h} \log_{10} n_{.j.h}$$

$$(15) \qquad \log_e \prod_{kh} n_{..kh}^{n_{..kh}} = 2.30258 \sum_{kh} n_{..kh} \log_{10} n_{..kh}$$

$$(16) \qquad \log_e n_{....}^{n_{....}} = (2.30258)(n_{....} \log_{10} n_{....})$$

$$(17) \qquad \log_e n_{....}^{2n_{....}} = (2.30258)(2n_{....} \log_{10} n_{....})$$

$$(18) \qquad \log_e n_{....}^{3n_{....}} = (2.30258)(3n_{....} \log_{10} n_{....})$$

Descriptions of the algebra necessary to test each of the six kinds of hypotheses along with appropriate examples are presented below.

a. Are the four attributes mutually independent?

The null hypothesis tested here is

H: $p_{ijkh} = p_i p_j p_k p_h$

The likelihood ratio is

$$\lambda = \frac{(\prod_i n_{i...}^{n_{i...}})(\prod_j n_{.j..}^{n_{.j..}})(\prod_k n_{..k.}^{n_{..k.}})(\prod_h n_{...h}^{n_{...h}})}{(n_{....}^{3n_{....}})(\prod_{ijkh} n_{ijkh}^{n_{ijkh}})}$$

with df = rstu - (r + s + t + u) + 3
and $\log_e \lambda$ = (1) + (2) + (3) + (4) -(18) - (5)

Chi-square for testing the above hypothesis is calculated by $X^2 = -2 \log_e \lambda$. The same procedure for calculating each chi-square from the appropriate λ was followed throughout.

b. Are three of the attributes mutually independent?

Four hypotheses may be tested here, namely those testing the mutual independence of A, B, and C; A, B, and D; B, C, and D; or A, C, and D. As an example, the hypothesis for testing the mutual independence of A, B, and C is

H: $p_{ijk} = p_i p_j p_k$

The associated contingency table may be thought of as a three-way table with cell frequencies $n_{ijk..}$. The likelihood ratio is

$$\lambda = \frac{(\prod_i n_{i...}^{n_{i...}})(\prod_j n_{.j..}^{n_{.j..}})(\prod_k n_{..k.}^{n_{..k.}})}{(\;{}^{2n} \qquad {}^{n_{ijk}}}$$

with df = rstu - (r + s + t) + 2
and $\log_e \lambda$ = (1) + (2) + (3) - (17) - (9)

c. Are two of the attributes independent?

Six hypotheses may be tested here, namely, those testing independence of A and B, A and C, A and D, B and C, B and D, and C and D. As an example, the hypothesis for testing the independence of A and B is

H: $p_{ij} = p_i p_j$

The likelihood ratio is

$$\lambda = \frac{(\prod_i n_{i...}^{n_{i...}})(\prod_j n_{.j..}^{n_{.j..}})}{(n_{....}^{n_{....}})(\prod_{ij} n_{ij..}^{n_{ij..}})}$$

with df = (r - 1)(s - 1)
and $\log_e \lambda$ = (1) + (2) - (16) - (10)

d. Is one attribute independent of a classification formed by combining the other three attributes into a unit?

Four hypotheses may be tested here, namely, those testing the independence of A from (B, C, D); B from (A, C, D); C from (A, B, D); and D from (A, B, C). As an example, the null hypothesis for the independence of A from (B, C, D) is

H: $p_{ijkh} = p_i p_{jkh}$

where p_{ijkh} is the probability of an individual's being in the ijkhth cell, p_i the probability of being in the ith category of attribute A and p_{jkh} the probability of being jointly in the jth, kth, and hth categories of attributes B, C, and D.

The likelihood ratio is

$$\lambda = \frac{(\Pi_{i} n_{i...}^{n_{i...}})(\Pi_{jkh} n_{.jkh}^{n_{.jkh}})}{(n_{....}^{n_{....}})(\Pi_{ijkh} n_{ijkh}^{n_{ijkh}})}$$

with df = $(r - 1)(stu - 1)$
and $\log_e \lambda = (1) + (6) - (16) - (5)$

e. Is one attribute independent of a classification formed by combining two of the other attributes into a unit?

Twelve hypotheses may be tested here, namely, those testing whether A is independent of either B and C, B and D, or C and D; whether B is independent of either A and C, A and D, or C and D; whether C is independent of either A and B, A and D, or B and D; and whether D is independent of either A and B, A and C, or B and C. As an example, the hypothesis for testing the independence of A with B and C is

H: $p_{ijk} = p_i p_{jk}$

where p_{ijk} is the probability of being in cells formed by summing over the h's and p_{jk} is the probability of being in the joint jth and kth categories of attributes B and C.

The likelihood ratio is

$$\lambda = \frac{(\Pi_{i} n_{i...}^{n_{i...}})(\Pi_{jk} n_{.jk.}^{n_{.jk.}})}{(n_{....}^{n_{....}})(\Pi_{ijk} n_{ij.h}^{n_{ij.h}})}$$

with df = $(r - 1)(st - 1)$
and $\log_e \lambda = (1) + (13) - (16) - (14)$

f. Is a classification formed by the combination of two attributes independent of a classification formed by the combination of the other two attributes?

Three hypotheses may be tested here, namely, those testing whether A and B is independent of C and D; whether A and C is independent of B and D; and whether A and D is independent of B and C. As an example, the hypothesis for testing the independence of A and B with B and C is

H: $p_{ijkh} = p_{ij} p_{kh}$

where p_{ij} is the probability of being in the ith and jth categories of attributes A and B jointly and p_{kh} the probability of being in the kth and hth categories of attributes C and D jointly.

The likelihood ratio is

$$\lambda = \frac{(\Pi_{ij} n_{ij..}^{n_{ij..}})(\Pi_{kh} n_{..kh}^{n_{..kh}})}{(n_{....}^{n_{....}})(\Pi_{ijkh} n_{ijkh}^{n_{ijkh}})}$$

with df = $(rs - 1)(tu - 1)$

The Exact Method of Freeman and Halton

The method of Freeman and Halton (5) is an extention of Fisher's exact method for the 2 × 2 case to tables with any number of attributes and any number of associated categories. It applies to any sampled array of numbers, whatever the population from which the sampling is effected, provided that: (a) either the parent population is infinite, or the sampling is done with replacement of the sampled members; and (b) the sampling is random.

Like Fisher's method, a special case of the present method for the 2 × 2 table, Freeman and Halton's method is applicable to tables which have theoretical frequencies so small that the chi-square test is not appropriate; even tables with some zero frequencies can be handled. According to the authors, the only difficulty in using the method is the amount of computational labor for dealing with large samples.

The method in general consists of assuming the border totals fixed, considering only relationships internal to the contingency table, considering every possible array of cell frequencies with the given border totals, and applying a test of significance as follows: (a) all arrays subject to the same general conditions as observed (i.e., the same border totals); (b) the corresponding a priori probabilities are calculated by means of the appropriate probability expression; (c) the values of the a priori probabilities smaller than or equal to the probability of the observed array are noted; these are the probabilities of all arrays which are a priori as probable as, or less probable than the observed array; (d) all probabilities satisfying the conditions in (c) are summed to yield the probability of obtaining an array as probable as or less probable than the observed array.

The general expression for the probability formula for any array stated in verbal terms is as follows: The probability of any array is equal to a ratio where the numerator is the product of the factorials of all border totals and the denominator is the product of the total sample size raised to the power of one less than the number of attributes and of the factorials of all cell frequencies.

The formula for computing the probability of having obtained any array through random sampling from a set of all arrays with fixed marginal totals is given symbolically for the two-way case as

$$P = \frac{\prod_i n_{i.}! \ \prod_j n_{.j}!}{n_{..}! \prod_i \prod_j n_{ij}!}$$

and for the three-way case as

$$P = \frac{\prod_i n_{i..}! \ \prod_j n_{.j.}! \ \prod_k n_{..k}!}{(n_{...})^2 \prod_i \prod_j \prod_k n_{ijk}!}$$

Design of the Survey

The Questions Under Investigation—The questions under investigation from which the null hypotheses were deduced were these:

1. Are attributes indicating the extent to which graduates utilized their training in actual teaching positions related to those attributes which were known about them at the time of their graduation?
2. Are the attributes which were known about graduates at the time of their graduation interrelated?
3. What were the reasons for graduates' not entering teaching or for failing to teach during one or more of the first five years?

Description of the Variables—The variables selected for study were considered as two sets of attributes: (1) pregraduation attributes, and (2) postgraduation attributes. The twelve pregraduation attributes were

1. Year graduated
2. Quarter graduated
3. Sex
4. Major field
5. Honors at commencement
6. Previous degree
7. Rank with major peer group
8. Age
9. Teaching previously
10. Duration of previous teaching
11. Military service
12. Extra-curricular activities

The four postgraduation attributes were

1. Taught since graduation
2. Taught during fifth year
3. Taught continuously
4. Intention toward future teaching

The attributes will be described later in some detail. First, however, the concept of "attribute" will be defined.

The term "attribute" in the present paper refers to a set of mutually and exhaustive "categories", so that any person in a sample of individuals can be classified as belonging in only one category of a given attribute. The simplest form of attribute is the dichotomy, and most of the attributes considered here are of this type. Thus, a person either has or does not have the characteristic specified by the attribute. For example, persons were classified on the basis of "honors" by the categories "yes" and "no". Placement in category "yes" means that the person had honors conferred at commencement, while "no" designates that no honors were conferred. The process of classifying persons into categories of an attribute involves the operation of counting the discrete number of persons in each category. This operation is fundamentally different from the process of measurement where a scale of more or less equal units is applied to a person or object and the measured number is an approximate number involving a whole number and a decimal fraction, such as in length the quantity 6.54 inches.

When a sample of individuals are classified on the basis of two attributes simultaneously, we count the number of persons who may possess jointly any of the possible combinations of pairs of categories. Again, it should be noted that classification is mutually exclusive and exhaustive. Thus, any person or object would be counted once and only once for any given contingency table.

The pregraduation and postgraduation attributes which were listed previously will now be described briefly.

The first of the pregraduation attributes was "year graduated." A year here refers to a period beginning with a December commencement and extending through an August commencement. The particular years given are the years during which the August commencements fell. This attribute was a dichotomy, and the two categories were

1948
1949

The second pregraduation attribute was "quarter graduated." This refers to the session during which the degree was conferred. The two summer sessions were treated as quarters. The five categories were

Fall
Winter
Spring
Summer Session I (July commencement)

Summer Session II (August commencement)

The two categories of the third attribute, "sex" were

Male
Female

The fourth pregraduation attribute was "major field." The fifteen categories were

Art education
Business education
Elementary education
English
Foreign languages (French, German, Spanish and Latin)
Industrial education
Mathematics
Music education
NKP (Nursery-kindergarten-primary)
Physical education (men)
Physical education (women)
Science
Social studies
Speech
Special fields (speech pathology and handicapped children)

The fifth pregraduation attribute was "honors at commencement." The two categories were

Honors conferred (Yes)
No honors conferred (No)

The sixth pregraduation attribute was "previous degree." This attribute refers to a graduate's having had a bachelor's degree previous to earning the B. S. in Education degree. The categories were

Had previous degree (Yes)
Had no previous degree (No)

The seventh pregraduation attribute was "rank with major peer group." The ranks referred to were obtained from rank-order lists submitted by the major departments to the Bureau of Recommendations each spring. The three categories were

High
Medium
Low

The eighth pregraduation attribute was "age at time of graduation." Initially, this was a continuous variable, and the age was determined to the nearest whole year. The range of ages was divided into four class-intervals. The four categories in terms of the limits of the class-intervals were

20-22 years
23-27 years
28-31 years
32-59 years

The four class-intervals for "age" were treated just as if they were categories of an attribute.

The ninth pregraduation attribute was "taught previously." The two categories were

Taught previously (Yes)
Had not taught previously (No)

The tenth pregraduation attribute was "duration of previous teaching." Like "age", this was initially a continuous variable, and duration was determined to the nearest whole year. The range of durations in years was divided into two class-intervals, so that the attribute became a dichotomy. The two categories were

1- 4 years
5-30 years

The eleventh pregraduation attribute was "military service." This refers only to military service during World War II or between the end of World War II and commencement. The two categories were

Veteran (Yes)
Non-veteran (No)

The twelfth pregraduation attribute was "extra-curricular activities." Evidence of participation in extra-curricular activities was obtained from an open-ended question on the Registration Blank of the Bureau of Recommendations. To be classified as a participant, a graduate must have shown at least four acceptable activities from school and college combined. Acceptable activities did not include jobs for which pay might have been received, activities in military service, academic scholarship or organizations, athletic activities for physical education majors or art activities for art majors. The two categories were

Evidence of marked participation (Yes)
No evidence (No)

The first of the postgraduation attributes was "taught since graduation." This attribute was a dichotomy in which possession of the attribute indicates that the graduate has done some teaching during the first five years since graduation. The two categories were

Taught since graduation (Yes)
Had not taught since graduation (No)

The second of the postgraduation attributes was "taught during fifth year." This attribute is a dichotomy in which possession of the attribute indicates that the graduate was teaching during the fifth year after graduation. The two categories were

Taught during fifth year (Yes)
Did not teach during fifth year (No)

The third postgraduation attribute was "taught continuously." This attribute was treated both as a trichotomy and as a dichotomy when either the latter two categories or the first and third categories were combined. The three categories were

Had taught continuously since graduation
Had teaching service interrupted for at least one year
Had not taught since graduation

The fourth postgraduation attribute was "intention toward teaching." This attribute was a polytomy with four categories, although in some cases categories were combined in order to facilitate the analysis. The categories were set up on the assumption that they formed a hierarchy of postgraduation experiences and attitudes in regard to the extent to which graduates actually intended to apply their education training in actual teaching positions. The four categories were

Teaching during fifth year
Not teaching; plan to teach in future
Not teaching; uncertain about future teaching
Not teaching; do not plan to teach in future

Population and Samples—The population was defined as graduates of the College of Education of the University of Minnesota. The two samples in the present study consisted of two classes, each class including graduates from all commencements over a one-year period extending from a fall commencement of one year through the August commencement of the following year.

Collection and Analysis of Data

Sources of Data—Pregraduation data were collected from personnel records in the Bureau of Recommendations of the University of Minnesota; postgraduation data were collected by means of a mail questionnaire. Pregraduation data were available for every graduate; in collecting postgraduation data, attempt was made by means of persistent follow-up procedures to secure at least a 90 percent response to the mail questionnaire.

The entire questionnaire was printed on the back side of a postal card. The instructions asked the respondent to indicate every school year in which he had taught since graduation by marking the year from a list on the card. For every year in which he had not taught, he was asked to indicate the reason or reasons for not teaching from a list of reasons given in the accompanying letter. The list of reasons was as follows:

1. Salary
2. Working conditions
3. Health (personal or family)
4. Maternity
5. Military service
6. Further study
7. Marriage
8. Held non-teaching job in business or industry
9. Suitable teaching jobs unavailable
10. Other reasons (to be specified by respondent)

Response to the Questionnaire—Of the 940 persons in the total samples for the two years, two were found to be deceased. Attempt was made to reach the remaining 938 cases by means of the mail questionnaire.

Table I presents the number and percentage of persons responding to the questionnaire, along with pertinent information about the extent to which persons could be reached. Based upon the estimated number of persons reached, the response to the questionnaire was 91.3 percent for the total sample.

Analysis by Simple Contingency Tables—In answering the first of the questions under investigation, which concerns the relationships between pregraduation and postgraduation attributes, analyses were made largely by means of simple, or two-way contingency tables; however, in some instances complex tables were employed. A thorough analysis was made of associations between the more than one hundred possible pairs of attributes which may be set up from the twelve pregraduation attributes and the four postgraduation attributes, where associations were, for the most part, studied separately for the two years of graduation. Since these analyses involved merely the usual chi-square test and comprise reiterative processes largely, a summary of the results of the complete set of analyses directed toward answering the first question under investigation is given in the present paper. Details of the tables have been reported elsewhere by the author (7). The results are summarized in Table II, where the probabilities of all significant chi-squares are tablulated.

It is appropriate here to specify the nature of the associations shown to be significant in Table II. The evidence here, which bears upon the first of the basic questions, showed nearly all of the pregraduation attributes to be associated with one or more of the postgraduation attributes.

The attribute "taught since graduation" was found to be associated more often with the graduates who were female, who were not majoring in social science, who were younger, and who

TABLE I

PARTICIPATION IN FOLLOW-UP STUDY OF PERSONS GRADUATING FROM THE COLLEGE OF EDUCATION OF THE UNIVERSITY OF MINNESOTA WITH THE B.S. IN EDUCATION DEGREE BETWEEN DECEMBER, 1947, AND AUGUST, 1949, INCLUSIVE, AND REGISTERING WITH THE BUREAU OF RECOMMENDATIONS

Category	Number	Per Cent Total N = 940	Per Cent Contacted N = 916
Total number of graduates during period	940	100.0%	- - - - -
Number of graduates who could not be contacted by mail questionnaire	22	2.3	- - - - -
Number of graduates known to be deceased	2	0.002	- - - - -
Number of graduates thought to have been contacted by mail questionnaire	916	97.7	100.0%
Number of graduates for whom questionnaires were completed	836	89.1	91.3
Number of graduates for whom questionnaires were not completed	80	8.5	8.7

*The total sample included all majors from the Minneapolis campus except Nursing Education, Recreation Leadership, and Library Science.

TABLE II

SUMMARY OF SIGNIFICANT ASSOCIATIONS BETWEEN PREGRADUATION AND POSTGRADUATION ATTRIBUTES AS SHOWN BY CHI-SQUARE TESTS IN SIMPLE CONTINGENCY TABLES FOR 1948 AND 1949 GRADUATES

		Postgraduation Attribute			
Pregraduation Attribute	Year of Graduation	Taught Since Graduation	Taught Fifth Year	Taught Continuously	Intention Toward Teaching
Year Graduated	1948				
	1949				
Quarter Graduated	(years combined)		.05		
Sex	1948		.001	.001	.01
	1949	.02	.001	.02	.05
Major Field	1948				
	1949	.01		.001	
Honors	1948			.05	.01
	1949				.01
Previous Degree	1948				
	1949				
Rank	1948				.001
	1949				
Age	1948		.01	.05	
	1949	.001	.001	.001	.01
Taught Previously	1948		.05	.05	
	1949		.001	.001	.001
Duration of Teaching	1948				
	1949			.02	
Military Service	1948		.01		.001
	1949	.02	.001	.02	.001
Extra-Curricular	1948		.001	.01	.001
	1949		.001	.05	.01

were not veterans.

The attribute "taught during fifth year" was found to be associated more often with the graduates who were summer graduates, who were male, who were older, who had taught previous to graduation, who were veterans, and who were participants in extra-curricular activities.

The attribute "taught continuously" was found to be associated more often with the graduates who were male, who were older, who had taught previous to graduation, who had taught for a longer duration previous to graduation, and who were veterans. Among all graduates those who had interrupted teaching service were found most often to be those who were honor graduates, who were non-veterans, and who participated in extra-curricular activities. Of the social science majors who did enter teaching, less had interrupted teaching service.

The attribute "intention toward teaching" was found to be related to pregraduation attributes as follows: the females either planned to teach in the future (if not now teaching) or else were uncertain about future plans to teach; those who were ranked low in their major peer group planned not to teach in the future more than those ranked medium or high; older graduates planned to teach in the future; those who had taught previously planned to teach in the future; those who had not taught previously were either uncertain or did not plan to teach in the future; non-veterans were uncertain about future plans to teach; participants in extra-curricular activities were either uncertain or planned to teach in the future.

Two simple tables will be considered in some detail, one to illustrate the exact method of Freeman and Halton, the other to illustrate scoring methods. The exact method is presented in this section of the paper, while scoring methods are deferred until after the illustrations on complex tables. The reason for this particular order of presentation is that the data for the scoring illustrations depends upon the results of analyzing the complex tables.

In order to illustrate Fisher's exact treatment for 2 × 2 contingency tables and the exact treatment of Freeman and Halton for contingency tables other than the 2 × 2, the association between "extra-curricular activities" and "taught fifth year" was investigated in a small sample of twenty cases. This sample consisted of the Industrial Education majors for the 1948 graduates. These data are given in Table III. The obtained chi-square by the uncorrected shortcut formula was 4.3, with 1 df, which is significant at the .05 level. Chi-square by the corrected shortcut formula was 2.1 with 1 df, which is not significant at the .05 level. These results leave one in something of a dilemma. If we accept the results of the corrected formula, we must conclude that the results are non-significant. If we accept the results of the uncorrected formula, we conclude that the results are significant, but with the knowledge that we have violated the rule for minimum-sized theoretical frequencies. Such a situation as this is one where exact methods are very applicable.

First, we use Fisher's exact method to determine whether the results are in fact significant or not. The principle of Fisher's exact test consists in assuming that the marginal totals remain fixed, using a formula yielding the probability of having any set of cell frequencies, calculating the probabilities for the observed set of cell frequencies and all more "extreme" arrays (i.e., arrays which would yield a chi-square of greater magnitude than that for the observed array), and summing these probabilities. This summed probability gives the probability of having obtained by chance alone an array as "extreme" or "more extreme" than the observed array, and this probability is then referred to some predetermined significance level, such as the .05 level. For example, if the obtained probability were more than .05, we would accept the null hypothesis, and we would reject the hypothesis if the probability were less than .05. It would be noted that this is a one-tailed test of significance.

Using the exact formula given previously for the two-way table, we find that the probability of the observed array as given by the data of Table III is

$$P = \frac{11!\ 9!\ 3!\ 17!}{20!\ 0!\ 11!\ 3!\ 6!} = 0.074$$

Since no other array with the same marginal totals has a probability smaller than or equal to 0.074, this is in fact the probability corresponding to the significance level for the array. Since this probability is exact, it is evident that the uncorrected shortcut formula gave an underestimate of the probability and the corrected shortcut formula an overestimate. The term "exact" here refers to the fact that an exact test leads to exact probabilities for testing significance.

It was suspected that a significant association might be found if the degrees of freedom for the sample of twenty cases were increased. Therefore, the data were rearranged into a 2 × 3 contingency table by a regrouping of the basic categories; the former category of "did not teach during fifth year" was divided into two new categories. The data are given in Table IV.

The exact probability of obtaining the array in Table IV by random sampling alone is given by

$$P = \frac{3!\ 17!\ 11!\ 7!\ 2!}{20!\ 0!\ 3!\ 0!\ 11!\ 4!\ 2!} = 0.030702$$

TABLE III

2 x 2 CONTINGENCY TABLE TO TEST ASSOCIATION OF "EXTRA-CURRICULAR ACTIVITIES" AND THE ATTRIBUTE, "TAUGHT DURING FIFTH YEAR", FOR 1948 INDUSTRIAL EDUCATION MAJORS

Extra-Curricular Activities	Taught During Fifth Year		
	Yes	No	
Yes	0 (0.0%)	3 (100.0%)	3 (100.0%)
No	11 (64.7%)	6 (35.3%)	17 (100.0%)
	11	9	20

$x^2 = 4.3$ (without Yates' correction)

$df = 1$

$.02 < P < .05$

$x^2 = 2.1$ (with Yates' correction)

$df = 1$

$.10 < P < .20$

TABLE IV

2 x 3 CONTINGENCY TABLE TO TEST ASSOCIATION OF "EXTRA-CURRICULAR ACTIVITIES" AND THE ATTRIBUTE, "TAUGHT DURING FIFTH YEAR", FOR 1948 INDUSTRIAL EDUCATION MAJORS

Taught During Fifth Year	Extra-Curricular Activities		
	Yes	No	
Taught fifth year	0 (0.0%)	11 (64.7%)	11
Taught, left, and have not returned	3 (100.0%)	4 (23.5%)	7
Had not taught since graduation	0 (0.0%)	2 (11.8%)	2
	3 (100.0%)	17 (100.0%)	20

TABLE V

ALL POSSIBLE ARRAYS WITH FIXED MARGINAL TOTALS AS GIVEN BY THE DATA OF TABLE IV

a

0	11
3	4
0	2

b

0	11
2	5
1	1

c

0	11
1	6
2	0

d

1	10
2	5
0	2

e

1	10
1	6
1	1

f

1	10
0	7
2	0

g

2	9
1	6
0	2

h

2	9
0	7
1	1

i

3	8
0	7
0	2

TABLE VI

PROBABILITIES FOR ALL POSSIBLE ARRAYS WITH MARGINAL TOTALS AS GIVEN BY OBSERVED 2 x 3 TABLE OF ATTRIBUTES, "EXTRA-CURRICULAR ACTIVITIES" AND "TAUGHT DURING FIFTH YEAR", FOR 1948 INDUSTRIAL EDUCATION MAJORS

Array	n_{11}	n_{12}	n_{13}	n_{21}	n_{22}	n_{23}	P
a	0	3	0	11	4	2	.030 702
b	0	2	1	11	5	1	.036 842
c	0	1	2	11	6	0	.006 140
d	1	2	0	10	5	2	.202 632
e	1	1	1	10	6	1	.135 088
f	1	0	2	10	7	0	.009 649
g	2	1	0	9	6	2	.337 719
h	2	0	1	9	7	1	.096 491
i	3	0	0	8	7	2	.144 737
							1.000 000

TABLE VII

2 x 2 x 2 CONTINGENCY TABLE OF ATTRIBUTES, "SEX", "EXTRA-CURRICULAR ACTIVITIES", AND "TAUGHT FIFTH YEAR", FOR 1948 AND 1949 GRADUATES COMBINED

		Sex						
		Male			Female			
		Taught Fifth Year			Taught Fifth Year			
		Yes	No		Yes	No		
Extra-Curricular Activities	Yes	79	62	141	73	168	241	382
	No	177	121	298	81	75	156	454
		256	183	439	154	243	397	836

Male:

$X^2 = 0.447$

df = 1

$P < .05$

Female:

$X^2 = 18.665$

df = 1

$P < .001$

The method of Freeman and Halton asks us to consider the probabilities of all possible arrays for a given set of marginal totals. Before presenting the probabilities for these arrays, it is well to consider the nature of the arrays themselves. The nine 2 × 3 tables given in Table V (a through i) represent all possible arrays with the marginal totals of Table IV. Array a of Table V is the same as the data of Table IV. Probabilities of the other eight arrays were computed in like manner. The probabilities for the complete set of nine arrays are given in Table VI. There are two probabilities smaller than that of the observed array a, which is 0.030702; these two probabilities are those for arrays c and f and are numerically equal to 0.006140 and 0.009649, respectively. When these three probabilities are summed, the total is equal to 0.046491. This figure represents the probability of having the observed array or any array more extreme. Since the obtained probability is less than .05, we reject the null hypothesis of no association between the attributes.

The results of the foregoing test of significance indicates that there is evidence of association between "extra-curricular activities" and "taught during fifth year" for the 1948 Industrial Education majors. If, however, one had used only the 2 × 2 table, the evidence would not have been found; a 2 × 3 table was needed in order to uncover the relation.

These results raise the question of the nature of the association. The association may be interpreted by noting the percentages in the various cells of the 2 × 3 table. Among the three persons who had participated in extra-curricular activities, all (or 100.0 percent) were found to have taught sometime during the first five years, to have left teaching and not to have returned by the fifth year. Among the seventeen persons who had not participated in extra-curricular activities, a relatively high percentage (64.7) were found to have taught during the fifth year. We may state the relationship, then, by saying that participants in extra-curricular activities tended not to teach during the fifth year, while non-participants tended, by and large, to teach during the fifth year. These results are similar to those found for all majors for the 1948 and 1949 samples. In addition, for the Industrial Education majors it was found that participants tended to teach for a short period, to leave teaching and not to return by the fifth year.

Analysis by Complex Contingency Tables—Two complex contingency tables are illustrated here. The first is a 2 × 2 × 2 table illustrating the method of Bartlett and Snedecor, while the other is a four-way table illustrating the likelihood ratio method.

In order to illustrate the method of Bartlett and Snedecor for the 2 × 2 × 2 table, the data were classified by the two pregraduation attributes "sex" and "extra-curricular activities," and by the postgraduation attribute, "taught during fifth year." The results of testing the association of the postgraduation attribute, "taught fifth year," with the pregraduation attributes, "sex" and "extra-curricular activities" individually in two-way tables had showed strong association (i.e., probabilities smaller than .001) in every instance. It was suspected that there might be association among all three of the attributes or what has been called "second-order interaction" by Bartlett and Snedecor. The kind of second-order interaction suspected in the present instance was that of a sex difference in the association between the two attributes, "extra-curricular activities" and "taught fifth year."

The 2 × 2 × 2 table is shown in Table VII. The first step in the analysis of the data was to examine each of the 2 × 2 tables which were classified by the two attributes, "extra-curricular activities" and "taught fifth year," for independence. Thus, there were separate 2 × 2 tables for the two sexes, one for males, the other for females. The obtained chi-square for males was 0.447, with 1 df, which is not significant at the .05 level; chi-square for the females was 18.665, with 1 df, which was significant at the .001 level.

The second order interaction was calculated as follows:

$$\frac{adb'c'}{a'd'bc} = \frac{(79)(121)(81)(168)}{(73)(75)(177)(62)} = 2.165$$

In order to test the hypothesis that in the population the second-order interaction is unity, the value of the deviation $|x|$ common to all eight cells of the 2 × 2 × 2 table was determined by solving the equation

$$(a-x)(d-x)(c'-x)(b'-x) = (c+x)(b+x)(a'+x)(d'+x)$$

Substituting the cell frequencies from our data, we obtain the equation

$$(79-x)(121-x)(168-x)(82-x) = (62+x)(176+x)(73+x)(75+x)$$

which reduces to

$$514\,x^3 - 24{,}664\,x^2 - 4{,}318{,}692\,x - 71{,}941{,}584 = 0$$

A first approximation may be obtained by ignoring the first two terms and solving the equation

$$4{,}318{,}692x - 71{,}941{,}584 = 0$$

Solving this equation, we obtain as a first approximation

$$x = \frac{71,941,584}{4,318,692} = 16.658$$

As a first step in finding two values of x between which the left-hand member of the equation changes sign, the obtained x above was rounded off to 16 and substituted in the cubic equation, yielding the value 5, 576, 816 for the left-hand member. The second step was to substitute the next lower integral number 15 for x, yielding the value 122, 946, which is smaller but which is not negative. When 14 was substituted for x, the left-hand member yielded the negative value -5, 235, 336, indicating the desired change in sign. Therefore, the value of x was known to lie between the values 14 and 15.

An improved approximation was obtained by solving the equation

$$\frac{x - 14}{15 - x} = \frac{5,235,336}{122,946}$$

where x = 14.9

A better approximation was desired, and it was found that the left-hand member was 14, 834.76 for 14.98 and was -39, 205.44 for x = 14.97. The equation

$$\frac{x - 14.97}{14.98 - x} = \frac{39,205.44}{14,834.76}$$

from which

$$x = 14.977$$

The obtained value of x, rounded off to two decimal places, or x = 14.98, was substituted in the formula for chi-square as follows:

$$X^2 = (|x| - 0.5)^2 \left(\frac{1}{a-x} + \frac{1}{d-x} + \frac{1}{c'-x} + \frac{1}{b'-x} + \frac{1}{c+x} + \frac{1}{b+x} + \frac{1}{a'+x} + \frac{1}{d'+x} \right)$$

$$= (14.98-0.5)^2 \left(\frac{1}{79-14.98} + \frac{1}{121-14.98} + \frac{1}{168-14.98} + \frac{1}{82-14.98} + \frac{1}{62+14.98} + \frac{1}{176+14.98} + \frac{1}{73+14.98} + \frac{1}{75+14.98} \right)$$

$$= (14.48)^2 \left(\frac{1}{64.02} + \frac{1}{106.02} + \frac{1}{153.02} + \frac{1}{67.02} + \frac{1}{76.98} + \frac{1}{190.98} + \frac{1}{87.98} + \frac{1}{89.98} \right)$$

$$= 224.400(0.0872)$$

$$= 19.568$$

The obtained chi-square for the second-order interaction is 19.568, with 1 df, which is significant at the .001 level. Therefore, we reject the null hypothesis of no interaction.

Rejecting the hypothesis of no interaction raises the question of the nature of the interaction. A significant interaction tells us that there is a difference in the nature of the association of the other two attributes for the sexes. The differences are such that among the males, participants and non-participants in extra-curricular activities were equally likely to have taught during fifth year, while among the females the participants in extra-curricular activities tended not to teach during the fifth year.

A four-way contingency table classified by "year", "rank", "honors" and "extra-curricular activities" is given in Table VIII. By means of the likelihood ratio, five hypotheses were tested for this table. They were

1. The four attributes, "year", "rank", "honors", and "extra-curricular activities" are mutually independent.
2. The three attributes, "rank", "honors", and "extra-curricular activities", are mutually independent.
3. The two attributes, "rank" and "honors" are independent.
4. The two attributes, "rank" and "extra-curricular activities" are independent.
5. The two attributes, "honors" and "extra-curricular activities" are independent.

In Table VIII the letters A, B, C, and D in our symbol system represent the attributes in the following order: "year", "rank", "honors", and "extra-curricular activities".

It will be recalled that any of the likelihood ratio criterion formulas for estimating chi-square for all possible hypotheses based upon a four-way table could be expressed in terms of combinations of eighteen terms which were natural logarithms of functions of cell or marginal totals. In order to test the five hypotheses for the present four-way table, only twelve of the eighteen terms were required. The twelve terms, along with the calculated numerical values are shown on the next page.

The required frequency totals for the twelve terms were obtained by summing frequencies according to the indicated process. For example, the totals denoted $n_{i...}$, which represent the total frequencies for the years 1948 and 1949, were obtained by adding all cell frequencies for

TABLE VIII

COMPLEX CONTINGENCY TABLE CLASSIFIED BY THE ATTRIBUTES, "YEAR", "RANK", "HONORS" AND "EXTRA-CURRICULAR ACTIVITIES" FOR GRADUATES OF 1948 AND 1949

		Honors						
		Yes			No			
		Extra-Curricular			Extra-Curricular			
Year	Rank	Yes	No		Yes	No		
1948	High	13	6	19	29	20	49	
	Medium	10	5	15	34	44	78	
	Low	3	1	4	35	34	69	
		26	12	38	98	98	196	234
1949	High	26	17	43	28	31	59	
	Medium	13	12	25	49	60	117	
	Low	6	5	11	45	51	96	
		45	34	79	122	150	272	351
$n_{..k.}$				117			468	

$n_{...h}$ 291
294

$n_{.j..}$ 170
235
180

$n_{....} = 585$

$$\log_e \prod_i n_{i...}^{n_{i...}} = 2.30258 \sum_i n_{i...} \log_{10} n_{i...} = 3332.67612 \tag{1}$$

$$\log_e \prod_j n_{.j..}^{n_{.j..}} = 2.30258 \sum_j n_{.j..} \log_{10} n_{.j..} = 3090.81427 \tag{2}$$

$$\log_e \prod_k n_{..k.}^{n_{..k.}} = 2.30258 \sum_k n_{..k.} \log_{10} n_{..k.} = 3434.65550 \tag{3}$$

$$\log_e \prod_h n_{...h}^{n_{...h}} = 2.30258 \sum_h n_{...h} \log_{10} n_{...h} = 3321.90198 \tag{4}$$

$$\log_e \prod_{ijkh} n_{ijkh}^{n_{ijkh}} = 2.30258 \sum_{ijkh} n_{ijkh} \log_{10} n_{ijkh} = 2030.79971 \tag{5}$$

$$\log_e \prod_{ijk} n_{ijk.}^{n_{ijk.}} = 2.30258 \sum_{ijk} n_{ijk.} \log_{10} n_{ijk.} = 2429.38732 \tag{9}$$

$$\log_e \prod_{jk} n_{.jk.}^{n_{.jk.}} = 2.30258 \sum_{jk} n_{.jk.} \log_{10} n_{.jk.} = 2820.43333 \tag{13}$$

$$\log_e \prod_{jh} n_{.j.h}^{n_{.j.h}} = 2.30258 \sum_{jh} n_{.j.h} \log_{10} n_{.j.h} = 2687.88973 \tag{14}$$

$$\log_e \prod_{kh} n_{..kh}^{n_{..kh}} = 2.30258 \sum_{kh} n_{..kh} \log_{10} n_{..kh} = 3032.68764 \tag{15}$$

$$\log_e n_{....}^{n_{....}} = (2.30258)(n_{....} \log_{10} n_{....}) = 3727.39025 \tag{16}$$

$$\log_e n_{....}^{2n_{....}} = (2.30258)(2n_{....} \log_{10} n_{....}) = 7454.78051 \tag{17}$$

$$\log_e n_{....}^{3n_{....}} = (2.30258)(3n_{....} \log_{10} n_{....}) = 11,182.17076 \tag{18}$$

1948 separately, then all cell frequencies for 1949. The summation process symbolically is

$$n_{i...} = \sum_h^u \sum_k^t \sum_j^s n_{ijkh}$$

and since 1948 has been taken arbitrarily as the first of the ith categories for "year", so that i = 1 for 1948, the total symbolically for 1948 would be

$$n_{1...} = \sum_h^u \sum_k^t \sum_j^s n_{1jkh}$$

This expression tells us to add all the cells for 1948 in a particular order, viz., to sum the categories for "rank" first, then "honors", then "extra-curricular activities". The numerical process was as follows: First, cell frequencies were added across the "rank" categories, so that in the upper left-hand 3 × 2 table we add for the first column

$$n_{1.11} = \sum_j^3 n_{1j11} = n_{1111} + n_{1211} + n_{1311} =$$

$$13 + 10 + 3 = 26$$

The three remaining totals for 1948 were added. These totals are numerically equal to 12, 98 and 98, respectively. The totals for extra-curricular activities were summed across the categories of honors as follows:

$$n_{1..h} = \sum_k^2 n_{1.kh}$$

The two totals for "extra-curricular activities" were

$$n_{1..1} = 26 + 98 = 124$$
$$n_{1..2} = 12 + 98 = 110$$

When these totals are summed across the two categories of "extra-curricular activities", we obtain the total frequency for 1948

$$n_{1...} = \sum_h^2 n_{1..h} = 124 + 110 = 234$$

The totals needed for the remaining eleven terms were obtained in similar fashion.

The twelve terms were calculated by using the summation form of the formula, by substituting logarithms to the base 10 initially, then changing to the base e by multiplying by the constant e = 2.30258. For example, term (1) was calculated numerically as follows:

$$\log_e \prod_i n_{i\ldots}^{n_{i\ldots}} = 2.30258 \sum_i n_{i\ldots} \log_{10} n_{i\ldots}$$

$$= 2.30258\ [(234)(2.36922) + (351)(2.54407)]$$

$$= (2.30258)(1447.36605)$$

$$= 3332.67612$$

In order to test the first hypothesis, that the four attributes are independent, we use the formula of the first kind of hypothesis for a four-way table. The expression for the likelihood criterion in terms of sums and differences of certain of the eighteen terms is as follows:

$$\log_e \lambda = (1) + (2) + (3) + (4) - (18) - (5)$$

Substituting the numerical values of the terms we have

$$\log_e \lambda = 3{,}332.67612 + 3{,}090.81427 + 3{,}434.65550 + 3{,}321.90198 - 11{,}182.17076 - 2{,}030.79971$$

$$= -32.92260$$

Using the formula for estimating chi-square in terms of λ, we have

$$X^2 = -2 \log_e \lambda$$

$$= (-2)(-32.92260)$$

$$= 65.84520$$

where the degrees of freedom is computed as follows:

$$df = rstu - (r + s + t + u) + 3$$

$$= (2)(3)(2)(2) - (2 + 3 + 2 + 2) + 3$$

$$= 18$$

The obtained chi-square for the first hypothesis, that is, 65.845, with 18 df, is significant at the .001 level. Therefore, we reject the null hypothesis.

In order to test the second hypothesis, that the three attributes "rank", "honors" and "extra-curricular activities" are mutually independent, we use a formula of the second type. The likelihood ratio criterion is

$$\log_e \lambda = (2) + (3) + (4) - (17) - (9)$$

Substituting the numerical values of the five terms, we have

$$\log_e \lambda = 3{,}090.81427 + 3{,}434.65550 + 3{,}321.90198 - 7{,}454.78051 - 2{,}429.38732$$

$$= -36.79608$$

from which

$$X^2 = 73.59216$$

The df is given by

$$df = rstu - (s + t + u) - 2$$
$$= 7$$

The obtained chi-square for the second hypothesis, that is, 73.592, with 7 df, is significant at the .001 level. Therefore, we reject the null hypothesis.

In order to test the third, fourth and fifth hypotheses, formulas of the third type were used. The appropriate formulas for an estimate of chi-square to test the third hypothesis, that "rank" and "honors" are independent, is expressed in terms of sums and differences of the eighteen terms as follows:

$$\log_e \lambda = 3{,}090.81427 + 3{,}424.65550 - 3{,}727.39025 - 2{,}820.43333$$

$$= -22.35381$$

from which

$$X^2 = 44.70762$$

The df is given by

$$df = (s - 1)(t - 1)$$
$$= 2$$

The obtained chi-square for the third hypothesis, that is, 44.708, with 2 df, is significant at the .001 level. Therefore, we reject the null hypothesis.

The appropriate formula for an estimate of chi-square to test the fourth hypothesis, that "rank" and "extra-curricular activities" are independent, is expressed in terms of certain of the eighteen terms as follows:

$$\log_e \lambda = (2) + (4) - (16) - (14)$$

Substituting the numerical values of the four terms, we obtain

$$\log_e \lambda = 3{,}090.81427 + 3{,}321.90198 - 3{,}727.39025$$

$$- 2,687.88973$$

$$= -2.56373$$

from which

$$X^2 = 5.12746$$

The df is given by

$$\begin{aligned} df &= (r - 1)(u - 1) \\ &= 2 \end{aligned}$$

The obtained chi-square for the fourth hypothesis, that is, 5.127, with 2 df, is not significant at the .05 level. Therefore, we do not reject the null hypothesis.

The appropriate formula for an estimate of chi-square to test the fifth hypothesis, that "honors" and "extra-curricular activities" are independent, is expressed in terms of sums and differences of certain of the eighteen terms as follows:

$$\log_e \lambda = (3) + (4) - (16) - (15)$$

Substituting the numerical values of the four terms, we obtain

$$\begin{aligned} \log_e \lambda &= 3,434.65550 + 3,321.90198 - 3,727.39025 \\ &\quad -3,032.68764 \\ &= -3.52041 \end{aligned}$$

from which

$$X^2 = 7.04082$$

The df is given by

$$\begin{aligned} df &= (t - 1)(u - 1) \\ &= 1 \end{aligned}$$

The obtained chi-square for the fifth hypothesis, that is, 7.041, with 1 df, is significant at the .01 level. Therefore, we reject the null hypothesis.

Analysis by Scoring Techniques—The methods of Yates and of Williams were illustrated with a 3 × 3 contingency table. The data are given in Table IX. Here the classification is based upon the three attributes, "rank", "honors", and "extra-curricular activities". The likelihood ratio analysis had shown mutual association among these three attributes. The classification of rows in Table IX is by categories of "rank", and the classification of columns is by a combination of the categories of "honors" and "extra-curricular activities".

The obtained chi-square for Table IX was 34.436, with 4 df, which is significant at the .001 level. Therefore, we reject the null hypothesis of independence of the classifications. Since this is a contingency table in which association is known to exist, the methods of Yates and Williams are applicable. Especially is this true since there appears to be a kind of departure in the table which resembles the arrangement of frequencies in a scatterdiagram in which positive correlation exists. For example, in the first column of Table IX there is a tendency for the cases to "pile up" in the lower left-hand corner (34.3 percent) as compared with those in the upper left-hand corner (20.6 percent), while in the third column there is a tendency for the cases to "pile up" in the upper right-hand corner (55.0 percent) as compared with those in the lower right-hand corner (12.7 percent). In the methods of Yates or of Williams, we test whether the association between row and column scores is significant, where the scores are assigned on the assumption of an underlying quantitative basis.

The Yates solution follows the representation in Figure 1. The numerical scoring scheme for rows and columns and the summation terms which were calculated with the aid of these scores are given in Table X. The scores for rows and columns were chosen arbitrarily. It was convenient in the present data to make the score for the middle category of the rows and columns each equal to zero, in order to simplify the calculations.

The three quantities, A, A' and B, needed for the regression formulas were calculated from the summation terms as follows:

$$\begin{aligned} A &= n_{..} \sum_i x_i^2 n_{i.} - (\sum_i x_i n_{i.})^2 \\ &= 585(350) - (-10)^2 \\ &= 204,650 \\ A' &= n_{..} \sum_j y_j n_{.j} - (\sum_j y_j n_{.j})^2 \\ &= 585(319) - (-177)^2 \\ &= 155,286 \\ B &= n_{..} \sum_i (x_i \sum_j y_j n_{ij}) - (\sum_i x_i n_{i.})(\sum_j y_j n_{.j}) \\ &= 585(64) - (-10)(-177) \\ &= 35,670 \end{aligned}$$

The two regression coefficients were calculated as follows:

$$b_{xy} = B/A$$

TABLE IX

CONTINGENCY TABLE TO TEST ASSOCIATION BETWEEN "RANK" AND COMBINATIONS OF "HONORS" AND "EXTRA-CURRICULAR ACTIVITIES"

	Combinations of Honors and Extra-Curricular Activities			
Rank	Both "no honors and "no extra-curricular"	Either "honors" and "no extra-curricular" or "no honors" and "extra-curricular"	Both "honors" and "extra-curricular"	
High	51 (20.6%)	80 (30.0%)	39 (55.0%)	170
Medium	112 (45.2%)	100 (37.6%)	23 (32.4%)	235
Low	85 (34.3%)	86 (32.3%)	9 (12.7%)	180
	248 (100.1%)	266 (99.9%)	71 (100.1%)	585

$x^2 = 34.436$

$df = 4$

$P < .001$

TABLE X

SCORING SCHEME AND SUMMATION TERMS IN YATES' SOLUTION OF DATA IN TABLE IX

	y_j^2	1	0	1		
	y_j	-1	0	+1		
x_i^2	x_i					$\sum_j y_j n_{ij}$
1	+1	51	80	39	170	-12
0	0	112	100	23	235	-89
1	-1	85	86	9	180	-76
		248	266	71	585	
$\sum_i x_i n_{ij}$		-34	-6	-30		

$\sum_i x_i n_{i.} = -10$ $\qquad \sum_i (x_i y_j n_{ij}) = 64$

$\sum_j y_j n_{.j} = -177$ $\qquad \sum_i (x_i^2 n_{i.}) = 350$

$\sum_j (y_j^2 n_{.j}) = 319$

$$= \frac{35,670}{204,650}$$

$$= 0.17430$$

and $$b_{yx} = B/A'$$

$$= \frac{35,670}{155,286}$$

$$= 0.22971$$

The standard errors of the regression coefficients were calculated as follows:

$$\text{S.E. of } b_{xy} = \sqrt{A'/An_{..}}$$

$$= \frac{155,286}{204,650(585)}$$

$$= 0.03601$$

and $$\text{S.E. of } b_{yx} = \sqrt{A/A'n_{..}}$$

$$= \frac{204,650}{155,286(585)}$$

$$= 0.04746$$

The significance of each of the regression coefficients was tested by means of the t test, where the hypothesis in each case was that the population regression coefficient was equal to zero. The t test for the b's were performed as follows:

$$t = \frac{b_{xy}}{\text{S.E. of } b_{xy}} = \frac{0.17430}{0.03601} = 4.840$$

$$t = \frac{b_{yx}}{\text{S.E. of } b_{yx}} = \frac{0.22971}{0.04746} = 4.840$$

Referring the obtained t's to the normal tables, we find that $P < .001$. Therefore, we reject the null hypothesis. We conclude, therefore, that the population regression coefficients are not equal to zero.

As an alternative test, we may square t and we have $X^2 = t^2 = (4.840)^2$

$$= 23.426$$

$$df = 1$$

$$P < .001$$

Therefore, the association of row and column scores is shown to be significant.

The row and column scores used in the Yates' method were arbitrary, and we do not know that they were the most appropriate for the data. The Williams' method provides empirically-derived scores which yield the maximum value of the correlation coefficient which would be possible for any set of scores for the same data. The Williams' solution follows the scheme given previously in this paper.

Starting with the original data in Table IX, the matrix G was computed by the formula for the elements as follows:

$$g_{ij} = n_{ij}/\sqrt{n_{i.}\,n_{.j}}$$

For example, the element of matrix G in the first row and first column was computed as follows:

$$g_{11} = n_{11}/\sqrt{n_{1.}\,n_{.1}}$$

$$= 51/\sqrt{(170)(248)}$$

$$= .24838$$

The matrix G is

$$G = \begin{pmatrix} .24838 & .37621 & .35500 \\ .46394 & .39997 & .17806 \\ .40231 & .39304 & .07961 \end{pmatrix}$$

The matrix T was obtained by $T = GG'$ and numerically the calculations were as given at the top of the next page.

In order to calculate the scores for rows and columns which would maximize the correlation between the two sets of scores, it was necessary, first, to solve for the second largest latent root λ of the matrix T. From T we have the determinant

$$\begin{vmatrix} .329\,2516 - \lambda & .328\,9174 & .276\,0529 \\ .328\,9174 & .406\,9217 - \lambda & .358\,0273 \\ .276\,0529 & .358\,0273 & .322\,6715 - \lambda \end{vmatrix}$$

which when expanded yields the characteristic equation

$$\lambda^3 - 1.0588\,\lambda^2 - 0.0589\,\lambda = 0$$

From Williams (12) we know that the largest root of the characteristic equation is unity, that the smallest root is zero, and that the remaining root, which we desire here, lies somewhere between zero and unity. Solving the cubic equation by the method previously given, we find that the desired root

$$\lambda = 0.0589$$

The value of this root gives the square of the maximized correlation between the variates assumed to underlie the row and column scores.

The correlation coefficient

$$T = \begin{pmatrix} .24838 & .37621 & .35500 \\ .46394 & .39997 & .17806 \\ .40231 & .39304 & .07961 \end{pmatrix} \begin{pmatrix} .24838 & .46394 & .40231 \\ .37621 & .39996 & .39304 \\ .35500 & .17806 & .07961 \end{pmatrix}$$

$$= \begin{pmatrix} .329\,2516 & .328\,9174 & .276\,0529 \\ .328\,9174 & .406\,9217 & .358\,0273 \\ .276\,0529 & .258\,0273 & .322\,6715 \end{pmatrix}$$

$$r_{xy} = \sqrt{\lambda} = 0.0589 = .24$$

The scores for rows of the contingency table were calculated by first solving for a set of solutions b_i to the set of simultaneous equations

$$[T - I]\,\{b_i\} = 0$$

and then finding a set of scores proportional to the b_i's so as to satisfy the restrictions

$$\sum_i n_{i.}\, x_i = 0$$

$$\sum_i n_{i.}\, x_i^2 = n_{..}$$

To solve for the b_i's we first subtract $\lambda=0.0589$ from the three main diagonal terms of the matrix T as follows:

$$T-\lambda I = \begin{pmatrix} .3293-.0589 & .3289 & .2761 \\ .3289 & .4069-.0589 & .3580 \\ .2761 & .3580 & .3227-.0589 \end{pmatrix}$$

$$= \begin{pmatrix} .2704 & .3289 & .2761 \\ .3289 & .3480 & .3580 \\ .2761 & .3580 & .2638 \end{pmatrix}$$

Using the elements of the matrix $T - \lambda I$ as coefficients, we write the system of simultaneous linear equations

$$\begin{aligned} .2704\, b_1 + .3289\, b_2 + .2761\, b_3 &= 0 \\ .3289\, b_1 + .3480\, b_2 + .3580\, b_3 &= 0 \\ .2761\, b_1 + .3580\, b_2 + .2630\, b_3 &= 0 \end{aligned}$$

Solving the equations, we obtain the values of the b_i's

$$\begin{aligned} b_1 &= 3.731 \\ b_2 &= -1.000 \\ b_3 &= -2.218 \end{aligned}$$

The set of b_i's were transformed to a set of row scores x_i so as to satisfy the restrictions given previously. The first step in this transformation was that of expressing the three row scores all in terms of one of the row scores. The score for the second row, x_2, is the easiest score computationally to deal with, for since we must obtain a set of x_i's proportional to the b_i's, we may express the three row scores in terms of x_2 as follows:

$$\begin{aligned} x_1 &= -3.731\, x_2 \\ x_2 &= x_2 \\ x_3 &= 2.218\, x_2 \end{aligned}$$

Substituting the values for the x_i's in terms of x_2 in the second restriction equation, we have

$$\sum_i n_i x_i^2 = 170(-3.731)^2 x_2^2 + 235\, x_2^2 + 180(2.218)^2 x_2^2 = 585$$

from which

$$\begin{aligned} x_1 &= 1.528 \\ x_2 &= -0.410 \\ x_3 &= -0.909 \end{aligned}$$

Checking to see that the obtained x_i's satisfy the two restrictions, we have

$$\sum_i n_{i.}\, x_i = 170(1.528) + 235(-0.410) + 180(-0.909) = -0.210$$

and $\sum_i n_{i.}\, x_i^2 = 170(1.528)^2 + 235(-0.410)^2 + 180(-0.909)^2 = 585.148$

and conclude that both restrictions are satisfied with negligible error.

The scores for columns of the contingency table were calculated as follows:

$$y_1 = \frac{51(1.528) + 112(-0.410) + 85(-0.909)}{(.24)(248)}$$

$$= -0.768$$

$$y_2 = \frac{80(1.528) + 100(-0.410) + 86(-0.909)}{(.24)(266)}$$

$$= 0.049$$

$$y_3 = \frac{39(1.528) + 23(-0.410) + 9(-0.909)}{(.24)(71)}$$

$$= 2.489$$

Checking to determine if the obtained y_j's satisfy the two restrictions, we have

$$\sum_j n_{.j} y_j = 248(-0.768) + 266(0.049) + 71(2.489) =$$

$$= -0.711$$

and $\sum_j n_{.j} y_j = 248(-0.768)^2 + 266(0.049)^2 + 71(2.489)^2 = 585.6$

and conclude that both restrictions are satisfied with negligible error.

As a check on the accuracy of the obtained row scores x_i and column scores y_j, the numerical values of the scores were substituted in the formula for the maximized correlation coefficient as follows:

$$r_{xy} = \Sigma\, n_{ij} x_i y_j / n_{..}$$

$$\begin{aligned} = [\, & 51(1.528)(-0.768) + 80(1.528)(0.049) \\ & + 39(1.528)(2.489) + 112(-0.410)(-0.768) \\ & + 100(-0.410)(0.049) + 23(-0.410)(2.489) \\ & + 85(-0.909)(-0.768) + 86(-0.909)(0.049) \\ & + 9(0.909)(2.489)] \,/585 = .24 \end{aligned}$$

which checks with the value of r_{xy} found by the expression

$$r_{xy} = \sqrt{\lambda}$$

The significance of the correlation was tested as follows:

$$X^2 = n_{..}\, r^2_{xy} = 585(.24)^2$$

$$= 33.696$$

$$df = 1$$

$$P < .001$$

Therefore, we conclude that the correlation is significant at the .001 level. According to Williams, "The interpretation of the data is simplest if the association can be accounted for satisfactorily in terms of the correlation between a single pair of variates corresponding to the two attributes." One can readily imagine how the attribute "rank" might have an underlying variate. Such a variate for the other attribute is not so readily conceived, although it is not inconceivable. One can only speculate as to the actual nature of the underlying variates, on the basis of the present data. We can interpret the association as being such that "low" ranking graduates tended not to be honor students and/or to be participants in extra-curricular activities, while, conversely, "high" ranking graduates tended also to be honor students and/or to be participants in extra-curricular activities.

Summary and Implications

This paper has described and illustrated a number of recent statistical techniques for analyzing association of attributes in contingency tables. Among the techniques are (1) those which permit the testing of hypotheses about regression and correlation of assumed underlying variates in r × s tables in which association is known to exist; (2) those which permit the testing of hypotheses involving three or four attributes simultaneously in an r × s × t × u contingency table; and (3) exact tests of significance for small sample data in r × s and r × s × t contingency tables.

The illustrative data were taken from an investigation of the postgraduate experiences of two graduating classes from a large teacher-training institution. Attributes of the graduates consisted of two main kinds: (1) pregraduation attributes, or those known at time of graduation, and (2) postgraduation attributes, or those characterizing the first five years of teaching experiences after graduation.

The statistical techniques presented here are applicable to many educational problems where they are not now being utilized extensively. In the first place, there are the instances where frequencies are often reported along with the corresponding percentages but without tests of significance, although generalizations are made from the data. Secondly, the availability of such techniques should suggest the use of previously ignored categorical data. It is recommended that such enumerative data augment rather than supplant continuous data. The techniques lend themselves well to data from surveys and to biographical data. Lastly, much data in the behavioral sciences are of such a nature that they are not immediately quantifiable; this does not mean, however, that they are inherently incapable of quantification. Often, our operation upon the data are of a primitive sort because of the immaturity of the science itself. It is to be hoped that techniques of contingency table analysis will assist in quantifying qualitative data.

It is also to be hoped that the techniques presented here will be given the widest possible dissemination and application to appropriate problems.

REFERENCES

1. Bartlett, M. S. "Contingency Table Interactions," Supplement to Journal of the Royal Statistical Society, II (1935), pp. 248-252.
2. Cochran, W. G. "Some Methods for Strengthening the Common X^2 Tests," Biometrics, X (1954), pp. 417-451.
3. Doi, A. N. Test for Independence of Four Classifications of a Complex Contingency Table by the Method of the Likelihood Ratio, unpublished M. A. colloquium paper, University of Minnesota, 1952, 44 pp.

4. Fisher, R. A. Statistical Methods for Research Workers, 10th Ed., Revised (New York: Hafner Publishing Co., 1946), 354 pp.
5. Freeman, G. H. and Halton, J. H. "Note on an Exact Treatment of Contingency, Goodness of Fit and Other Problems of Significance," Biometrika, XXXVIII (1951), pp. 141-149.
6. Gulliksen, H. Theory of Mental Tests (New York: John Wiley and Sons, 1950), 486 pp.
7. Mayo, S. T. Some Designs for the Collection of and Methods for the Analysis of Enumerative Data Characteristic of Attributes, with Special Applications to a Follow-up of Education Graduates. Unpublished Ph. D. Thesis (University Microfilm: Mic 56-2926), University of Minnesota, 1956, 283 pp.
8. Mood, A. McF. Introduction to the Theory of Statistics (New York: McGraw-Hill Book Co., 1950), 433 pp.
9. Norton, H. W. "Calculation of Chi-square for Complex Contingency Tables," Journal of the American Statistical Association, XL (1945), pp. 251-258.
10. Snedecor, G. W. Statistical Methods (Cedar Falls, Iowa: Iowa State College Press, 1946), 485 pp.
11. Stuart, A. "The Estimation and Comparison of Strengths of Association in Contingency Tables," Biometrika, XL (1953), pp. 105-110.
12. Williams, E. J. "Use of Scores for the Analysis of Association in Contingency Tables," Biometrika, XXXIX (1952), pp. 274-289.
13. Yates, F. "The Analysis of Contingency Tables with Groupings Based on Quantitative Characterters," Biometrika, XXXV (1948), pp. 176-181.

THE BEST LINEAR ESTIMATE OF THE PREDICTED VALUE AND THE STANDARD ERROR OF THE ESTIMATE

PALMER O. JOHNSON
University of Minnesota

DAVID AND NEYMAN* used the Markoff Theorem to derive a general method for solving the problem of obtaining the best linear estimate of a criterion and the standard error of the estimate. They gave as an illustration of the application of the theorem the case of one independent and one dependent variable, say X and Y. For this case there is little difficulty in calculations even with raw scores. For more than one independent variable, say $X_1, X_2, \ldots, X_s$, however, we suggest the use of deviation scores to save much computing work in deriving formulas for the problem we consider here. Our purpose is to obtain a general form for the estimation of Y, from a fixed set of more than one independent variable, i.e., from $X_1, X_2, \ldots, X_s$, and the determination of the standard error of such an estimate.

There are two situations that need to be differentiated with respect to the predicted value:

1. The estimate of the true mean Y-score, say, the average of all predicted $\tilde{y}$'s, or $E(\tilde{y})$'s, of all individuals in the population who have fixed, identical values of $X_1, X_2, \ldots, X_s$.
2. The true Y score for any single predicted $\tilde{y}$ score.

It is situation (1) with which this paper deals. We first summarize the notations and formulas used by David and Neyman. Several changes were made to fit our problem in general without distracting from the original meanings.

Define,

$$\left.\begin{array}{ll} Y_i & = \text{the Y score of the } i^{th} \text{ individual} \\ X_{1i} & = \text{the } X_1 \text{ score of the } i^{th} \text{ individual} \\ \vdots & \\ X_{si} & = \text{the } X_s \text{ score of the } i^{th} \text{ individual} \end{array}\right\} \quad (1)$$

When $i = 1, 2, \ldots n$; n denotes the number of individuals. Assume also that $X_1, X_2, \ldots, X_s$ are independent variables and that the relationship between Y and each X is linear.

Define again,

$$\left.\begin{array}{lll} y_i & = Y_i & - \overline{Y} \\ x_{1i} & = X_{1i} & - \overline{X}_1 \\ x_{2i} & = X_{2i} & - \overline{X}_2 \\ \ldots & \ldots & \ldots \\ x_{si} & = X_{si} & - \overline{X}_s \end{array}\right\} \quad (2)$$

When $y_i, x_{1i}, x_{2i}, \ldots, x_{si}$ are deviation scores of $Y, X_1, X_2, \ldots, X_s$, respectively, for the ith individual, $\overline{Y}, \overline{X}_1, \overline{X}_2, \ldots, \overline{X}_s$ are mean scores of $Y, X_1, X_2, \ldots, X_s$, respectively. With these conditions, we may write

$$\tilde{y}_i = E(\tilde{y}_i) = a_{i1}p_1 + \ldots + a_{is}p_s + a_{i(s+1)}\,p_{(s+1)} \quad (3)$$

Where $\tilde{y}_i$ or $E(\tilde{y}_i)$ denotes the mathematical expectation of y_i; p_1 and p_s are to be considered as unknown parameters. For our special case,

$$\left.\begin{array}{ll} a_{i1} & = 1 \\ a_{i2} & = x_1 \\ \ldots & \ldots \\ a_{is} & = x_{s-1} \\ a_{i(s+1)} & = x_s \end{array}\right\} \quad (4)$$

By using the Markoff Theorem, we obtain $\tilde{y}_i =$

$$E(\tilde{y}_i) = \frac{-\Delta_\theta}{\Delta} \quad (5)$$

$$\mu^2_{\tilde{y}_i} = -\frac{\Delta_0 \ \Delta_1}{(n-s-1)\,\Delta^2} \quad (6)$$

when $\mu^2_{\tilde{y}_i}$ denotes the square of the standard error for the estimate of each individual; n denotes the number of individuals involved, s is the number of independent variables. The terms Δ, Δ_θ, Δ_0, and Δ_1 are the determinants as indicated on the next page.

---

*F. N. David and J. Neyman. "Extension of the Markoff Theorem on Least Squares," Statistical Research Memoirs, II (December 1938), pp. 105-106.

$$\triangle = \begin{vmatrix} G_{11} & G_{12} & \cdots & G_{1s} & G_{1(s+1)} \\ G_{21} & G_{22} & \cdots & G_{2s} & G_{2(s+1)} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ G_{s1} & G_{s2} & \cdots & G_{ss} & G_{s(s+1)} \\ G_{(s+1)1} & G_{(s+1)2} & \cdots & G_{(s+1)s} & G_{(s+1)(s+1)} \end{vmatrix} \tag{7}$$

where $G_{jk} = \sum_{i=1}^{n} p_i \, a_{ij} \, a_{ik}$ (8)

For our problem, $p_i = 1$ (9)

$$\triangle_\theta = \begin{vmatrix} 0 & b_1 & b_2 & \cdots & b_s & b_{(s+1)} \\ H_1 & G_{11} & G_{12} & \cdots & G_{1s} & G_{1(s+1)} \\ H_2 & G_{21} & G_{22} & \cdots & G_{2s} & G_{2(s+1)} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ H_s & G_{s1} & G_{s2} & \cdots & G_{ss} & G_{s(s+1)} \\ H_{(s+1)} & G_{(s+1)1} & G_{(s+1)2} & \cdots & G_{(s+1)(s+1)} & G_{(s+1)(s+1)} \end{vmatrix} \tag{10}$$

where
$b_1 = 1$
$b_2 = x_1$
$b_3 = x_2$
. . . .
$b_s = x_{(s-1)}$
$b_{(s+1)} = x_s$ (11)

$H_j = \sum_{i=1}^{n} p_i \, y_i \, a_{ij}$ (12)

For our problem, $p_i = 1$ (13)

$$\Delta_0 = \begin{vmatrix} H_O & H_1 & H_2 & \cdots & H_S & H_{(s+1)} \\ H_1 & G_{11} & G_{12} & \cdots & G_{1S} & G_{1(s+1)} \\ H_2 & G_{21} & G_{22} & \cdots & G_{2S} & G_{2(s+1)} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ H_S & G_{S1} & G_{S2} & \cdots & G_{SS} & G_{S(s+1)} \\ H_{(s+1)} & G_{(s+1)1} & G_{(s+1)2} & \cdots & G_{(s+1)s} & G_{(s+1)(s+1)} \end{vmatrix} \tag{14}$$

$$\text{where } H_O = \sum_{i=1}^{n} \rho_i \, y_i^2 \tag{15}$$

$$\text{For our problem,} \qquad p_i = 1 \tag{16}$$

All the other notations are the same as indicated in (7) and (10).

$$\Delta_1 = \begin{vmatrix} 0 & b_1 & b_2 & \cdots & b_S & b_{(s+1)} \\ b_1 & G_{11} & G_{12} & \cdots & G_{1S} & G_{1(s+1)} \\ b_2 & G_{21} & G_{22} & \cdots & G_{2S} & G_{2(s+1)} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ b_S & G_{S1} & G_{S(2)} & \cdots & G_{SS} & G_{S(s+1)} \\ b_{(s+1)} & G_{(s+1)1} & G_{(s+1)2} & \cdots & G_{(s+1)s} & G_{(s+1)(s+1)} \end{vmatrix} \tag{17}$$

where all the notations are the same as indicated in (7) and (10).

Before we start to work out our problem, we give some elementary formulas which will be useful in writing our solution in more general form:

$$\left.\begin{array}{lll} r_{12.3} = \dfrac{r_{12} - r_{13} - r_{23}}{\sqrt{1-r_{13}^2}\ \sqrt{1-r_{23}^2}} & r_{1y.2} = \dfrac{r_{1y}-r_{12}r_{2y}}{\sqrt{1-r_{12}^2}\,\sqrt{1-r_{2y}^2}} & r_{1y.23} = \dfrac{r_{1y.3}-r_{12.3}r_{2y.3}}{\sqrt{1-r_{12.3}^2}\ \sqrt{1-r_{2y.3}^2}} \\ 1-R_{1.23}^2 = (1-r_{12}^2)(1-r_{13.2}^2) & 1-R_{y.12}^2 = (1-r_{1y}^2)(1-r_{2y.1}^2) & \\ 1-R_{y.123}^2 = (1-r_{1y}^2)(1-r_{2y.1}^2)(1-r_{3y.12}^2) & & \\ & \sigma_{y.1}^2 = \sigma_y^2(1-r_{1y}^2) & \\ & \sigma_{y.12}^2 = \sigma_y^2(1-R_{y.12}^2) & \\ \sigma_{1.2}^2 = \sigma_1^2(1-r_{12}^2) & & \\ \sigma_{1.23}^2 = \sigma_1^2(1-R_{1.23}^2) & & \\ \sigma_{y.123}^2 = \sigma_y^2(1-R_{y.123}^2) & & \end{array}\right\} \tag{18}$$

where 1, 2, 3 denotes x_1, x_2, x_3.

Now we consider the estimate of Y from only one variable X. By (1), (2), (8), (11), and (15), we have:

$$\left.\begin{array}{l} G_{11} = n,\ G_{12} = G_{21} = \Sigma x = 0, = G_{22} = \Sigma x^2 = n\sigma_x^2 \\ b_1 = 1,\ b_2 = x \\ H_0 = \Sigma y^2 = n\sigma_y^2,\ H_1 = \Sigma y = 0,\ H_2 = \Sigma xy = nr_{xy}\sigma_x\sigma_y \end{array}\right\} \tag{19}$$

From (7), (10), (14), and (17), we have

$$\triangle = \begin{vmatrix} n & 0 \\ 0 & n\sigma_x^2 \end{vmatrix} = n^2\sigma_x^2 \tag{20}$$

$$\triangle_\theta = \begin{vmatrix} 0 & 1 & x \\ 0 & n & 0 \\ nr\sigma_x\sigma_y & 0 & n\sigma_x^2 \end{vmatrix} = -n^2 r\sigma_x\sigma_y x \tag{21}$$

$$\triangle_0 = \begin{vmatrix} n\sigma_y^2 & 0 & nr\sigma_x\sigma_y \\ 0 & n & 0 \\ nr\sigma_x\sigma_y & 0 & n\sigma_x^2 \end{vmatrix} = n^3\sigma_x^2\sigma_y^2(1-r_{xy}^2) \tag{22}$$

$$\triangle_1 = \begin{vmatrix} 0 & 1 & x \\ 1 & n & 0 \\ x & 0 & n\sigma_x^2 \end{vmatrix} = -n(\sigma_x^2 + x^2) \tag{23}$$

From (5) we obtain: $\tilde{y} = r_{xy}\dfrac{\sigma_y}{\sigma_x}x$ (24)

If we want to estimate Y from X we simply use definition (2) and easily get

$$\tilde{y} = \overline{y} + r_{xy}\frac{\sigma_y}{\sigma_x}(X - \overline{X}) \tag{24a}$$

From (6) we obtain*

$$\mu_{\tilde{y}}^2 = \frac{\sigma_y^2(1-r_{xy}^2)}{n-2}\left\{1 + \frac{(X-\overline{X})^2}{\sigma_x^2}\right\} \quad \text{or} \quad \mu_{\tilde{y}}^2 = \frac{\sigma_{y.x}}{n-2}\left\{1 + \frac{(X-\overline{X})^2}{\sigma_x^2}\right\} \tag{25}$$

Next we consider the estimate of Y from two independent variables X_1 and X_2 .

---

* By elementary statistical knowledge it is not difficult to show that $\mu_{\tilde{Y}}^2 = \mu_{\tilde{y}}^2$. Hence this formula can also be used to evaluate the standard error of estimate for Y score. This holds when similar situations arise in this paper.

Following the same procedures, we have**

$$\Delta = \begin{vmatrix} n & 0 & 0 \\ 0 & n\sigma_1^2 & nr_{12}\sigma_1\sigma_2 \\ 0 & nr_{12}\sigma_1\sigma_2 & n\sigma_2^2 \end{vmatrix} = n^3\sigma_1^2\sigma_2^2(1-r_{12}^2) \tag{26}$$

$$\Delta_\theta = \begin{vmatrix} 0 & 1 & x_1 & x_2 \\ 0 & n & 0 & 0 \\ nr_{1y}\sigma_1\sigma_y & 0 & n\sigma_1^2 & nr_{12}\sigma_1\sigma_2 \\ nr_{2y}\sigma_2\sigma_y & 0 & nr_{12}\sigma_1\sigma_2 & n\sigma_2^2 \end{vmatrix} = -n^3\sigma_1\sigma_2\sigma_y\,[\,\sigma_2x_1(r_{1y}-r_{12}r_{2y})+\sigma_1x_2(r_{2y}-r_{12}r_{1y})] \tag{27}$$

$$\Delta_0 = \begin{vmatrix} n\sigma_y^2 & 0 & nr_{1y}\sigma_1\sigma_y & nr_{2y}\sigma_2\sigma_{2y} \\ 0 & n & 0 & 0 \\ nr_{1y}\sigma_1\sigma_y & 0 & n\sigma_1^2 & nr_{12}\sigma_1\sigma_2 \\ nr_{2y}\sigma_2\sigma_y & 0 & nr_{12}\sigma_1\sigma_2 & n\sigma_2^2 \end{vmatrix} = n^4\sigma_1^2\sigma_2^2\sigma_y^2\left\{(1-r_{12}^2)-r_1^2y-r_2^2y+2r_{12}r_{1y}r_{2y}\right\} \tag{28}$$

$$\Delta_1 = \begin{vmatrix} 0 & 1 & x_1 & x_2 \\ 1 & n & 0 & 0 \\ x_1 & 0 & n\sigma_1^2 & nr_{12}\sigma_1\sigma_2 \\ x_2 & 0 & nr_{12}\sigma_1\sigma_2 & n\sigma_2^2 \end{vmatrix} = -n^2\left\{(1-r_{12}^2)+\sigma_2^2x_1+\sigma_1^2x_2-2r_{12}\sigma_1\sigma_2x_1x_2\right\} \tag{29}$$

substituting the values of Δ, Δ_θ, Δ_0, and Δ_1 into (5) and (6), we obtain, by using formulas in (18):

$$\left.\begin{aligned} \tilde{y} &= r_{1y.2}\frac{\sigma_{y.2}}{\sigma_{1.2}}\,x_1 + r_{2y.1}\frac{\sigma_{y.1}}{\sigma_{2.1}}\,x_2 \\ \text{or } \tilde{Y} &= \overline{y} + r_{1y.2}\frac{\sigma_{y.2}}{\sigma_{1.2}}(X_1-\overline{X}_1) + r_{2y.2}\frac{\sigma_{y.1}}{\sigma_{2.1}}(X_2-\overline{X}_2) \end{aligned}\right\} \tag{30}$$

$$\begin{aligned} \mu_{F(y)}^2 &= \frac{\sigma_y^2(1-R_{y.12}^2)}{n-3}\left\{1+\frac{1}{1-r_{12}^2}\left[\frac{(X_1-\overline{X}_1)^2}{\sigma_1^2}+\frac{(X_2-\overline{X}_2)^2}{\sigma_2^2}-2r_{12}\frac{(X_1-\overline{X}_1)(X_2-\overline{X}_2)}{\sigma_1\sigma_2}\right]\right\} \\ \text{or } \mu_{F(y)}^2 &= \frac{\sigma_{y.12}^2}{n-3}\left\{1+\left[\frac{(X_1-\overline{X}_1)^2}{\sigma_{1.2}^2}+\frac{(X_2-\overline{X}_2)^2}{\sigma_{2.1}^2}-2r_{12}\frac{(X_1-\overline{X}_1)(X_2-\overline{X}_2)}{\sigma_{1.2}\,\sigma_{2.1}}\right]\right\} \end{aligned} \tag{31}$$

Then we consider the estimate of Y from three independent variables X_1, X_2, X_3.

**For simplicity, we use subscripts 1, 2, 3, ..., s to denote x_1, x_2, x_3, ..., x_s in writing r and σ. This holds throughout the paper.

Following the method used before, we have

$$\Delta = \begin{vmatrix} n & 0 & 0 & 0 \\ 0 & n\sigma_1^2 & nr_{12}\sigma_1\sigma_2 & nr_{13}\sigma_1\sigma_3 \\ 0 & nr_{12}\sigma_1\sigma_2 & n\sigma_2^2 & nr_{23}\sigma_2\sigma_3 \\ 0 & nr_{13}\sigma_1\sigma_3 & nr_{23}\sigma_2\sigma_3 & n\sigma_3^2 \end{vmatrix} = n^4\sigma_1^2\sigma_2^2\sigma_3^2\left[1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}r_{13}r_{23}\right] \tag{32}$$

$$\Delta_\theta = \begin{vmatrix} 0 & 1 & x_1 & x_2 & x_3 \\ 0 & n & 0 & 0 & 0 \\ nr_{1y}\sigma_1\sigma_y & 0 & n\sigma_1^2 & nr_{12}\sigma_1\sigma_2 & nr_{13}\sigma_1\sigma_3 \\ nr_{2y}\sigma_2\sigma_y & 0 & nr_{12}\sigma_1\sigma_2 & n\sigma_2^2 & nr_{23}\sigma_2\sigma_3 \\ nr_{3y}\sigma_3\sigma_y & 0 & nr_{13}\sigma_1\sigma_3 & nr_{23}\sigma_2\sigma_3 & n\sigma_3^2 \end{vmatrix}$$

$$= -n^4\sigma_1\sigma_2\sigma_3\left[\begin{array}{l} \sigma_2\sigma_3x_1\left\{r_{1y}(1-r_{23}^2)-r_{2y}(r_{12}-r_{13}r_{23})-r_{3y}(r_{13}-r_{12}r_{23})\right\} \\ +\sigma_1\sigma_3x_2\left\{r_{2y}(1-r_{13}^2)-r_{1y}(r_{12}-r_{13}r_{23})-r_{3y}(r_{23}-r_{12}r_{13})\right\} \\ +\sigma_1\sigma_2x_3\left\{r_{3y}(1-r_{12}^2)-r_{1y}(r_{13}-r_{12}r_{23})-r_{2y}(r_{23}-r_{12}r_{13})\right\} \end{array}\right] \tag{33}$$

$$\Delta_0 = \begin{vmatrix} n\sigma_y^2 & 0 & nr_{1y}\sigma_1\sigma_y & nr_{2y}\sigma_2\sigma_y & nr_{3y}\sigma_3\sigma_y \\ 0 & n & 0 & 0 & 0 \\ nr_{1y}\sigma_1\sigma_y & 0 & n\sigma_1^2 & nr_{12}\sigma_1\sigma_2 & nr_{13}\sigma_1\sigma_3 \\ nr_{2y}\sigma_2\sigma_y & 0 & nr_{12}\sigma_1\sigma_2 & n\sigma_2^2 & nr_{23}\sigma_2\sigma_3 \\ nr_{3y}\sigma_3\sigma_y & 0 & r_{13}\sigma_1\sigma_3 & nr_{23}\sigma_2\sigma_3 & n\sigma_3^2 \end{vmatrix}$$

$$= n^5\sigma_1^2\sigma_2^2\sigma_3^2\sigma_y^2\left[\begin{array}{l} (1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}r_{13}r_{23}) \\ -r_{1y}^2(1-r_{23}^2)-r_{2y}^2(1-r_{13}^2)-r_{3y}^2(1-r_{12}^2) \\ +2r_{1y}r_{2y}(r_{12}-r_{13}r_{23})+2r_{1y}r_{3y}(r_{13}-r_{12}r_{23}) \\ \qquad\qquad\qquad\qquad +2r_{2y}r_{3y}(r_{23}-r_{12}r_{13}) \end{array}\right] \tag{34}$$

$$\Delta_1 = \begin{vmatrix} 0 & 1 & x_1 & x_2 & x_3 \\ 1 & n & 0 & 0 & 0 \\ x_1 & 0 & n\sigma_1^2 & nr_{12}\sigma_1\sigma_2 & nr_{13}\sigma_1\sigma_3 \\ x_2 & 0 & nr_{12}\sigma_1\sigma_2 & n\sigma_2^2 & nr_{23}\sigma_2\sigma_3 \\ x_3 & 0 & nr_{13}\sigma_1\sigma_3 & nr_{23}\sigma_2\sigma_3 & n\sigma_3^2 \end{vmatrix} \tag{35}$$

(continued next page)

$$= -n^3 \begin{bmatrix} (1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}r_{13}r_{23}) + \sigma_2^2\sigma_3^2x_1^2(1-r_{23}^2) + \sigma_1^2\sigma_3^2x_2^2(1-r_{13}^2) \\ + \sigma_1^2\sigma_2^2x_3^2(1-r_{12}^2) - 2\sigma_1\sigma_2\sigma_3^2x_1x_2(r_{12}-r_{13}r_{23}) - 2\sigma_1\sigma_2^2\sigma_3x_1x_3(r_{13}-r_{12}r_{23}) \\ - 2\sigma_1^2\sigma_2\sigma_3x_2x_3(r_{23}-r_{12}r_{13}) \end{bmatrix} \tag{35}$$

After more troublesome algebraic manipulations, we obtain:

$$\left.\begin{aligned} \tilde{y} &= r_{1y.23}\frac{\sigma_{y.23}}{\sigma_{1.23}} + r_{2y.13}\frac{\sigma_{y.13}}{\sigma_{2.13}}x_2 + r_{3y.12}\frac{\sigma_{y.12}}{\sigma_{3.12}}x_3 \\ \text{or } \overline{Y} &= \overline{y} + r_{1y.23}\frac{\sigma_{y.23}}{\sigma_{1.23}}(X_1-\overline{X}_1) + r_{2y.13}\frac{\sigma_{y.13}}{\sigma_{2.13}}(X_2-\overline{X}_2) + r_{3y.12}\frac{\sigma_{y.12}}{\sigma_{3.12}}(X_3-\overline{X}_3) \end{aligned}\right\} \tag{36}$$

$$\mu_{\tilde{y}}^2 = \frac{\sigma_y^2(1-R_{y.123}^2)}{n-4}\left[1 + \frac{1}{1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}r_{13}r_{23}}\left\{\begin{aligned} &\frac{1-r_{23}^2}{\sigma_1^2}(X_1-\overline{X}_1)^2 \quad \frac{1-r_{13}^2}{\sigma_2^2}(X_2-\overline{X}_2)^2 \\ &+ \frac{1-r_{12}^2}{\sigma_3^2}(X_3-\overline{X}_3)^2 \\ &- \frac{2(r_{12}-r_{13}r_{23})}{\sigma_1\sigma_2}(X_1-\overline{X}_1)(X_2-\overline{X}_2) \\ &- \frac{2(r_{13}-r_{12}r_{23})}{\sigma_1\sigma_3}(X_1-\overline{X}_1)(X_3-\overline{X}_3) \\ &- \frac{2(r_{23}-r_{12}r_{13})}{\sigma_2\sigma_3}(X_2-\overline{X}_2)(X_3-\overline{X}_3) \end{aligned}\right\}\right]$$

$$\text{or } \mu_{\tilde{y}}^2 = \frac{\sigma_{y.123}^2}{n-4}\left[\begin{aligned} &1 + \frac{(X_1-\overline{X}_1)^2}{\sigma_{1.23}^2} + \frac{(X_2-\overline{X}_2)^2}{\sigma_{2.13}^2} + \frac{(X_3-\overline{X}_3)^2}{\sigma_{3.12}^2} - 2r_{13.3}\frac{(X_1-\overline{X}_1)(X_2-\overline{X}_2)}{\sigma_{1.23}\quad\sigma_{2.13}} \\ &- 2r_{13.2}\frac{(X_1-\overline{X}_1)}{\sigma_{1.23}}\frac{(X_3-\overline{X}_3)}{\sigma_{3.12}} - 2r_{23.1}\frac{(X_2-\overline{X}_2)(X_3-\overline{X}_3)}{\sigma_{2.13}\quad\sigma_{3.12}} \end{aligned}\right] \tag{37}$$

From all the above, we can finally derive the formulas for $\tilde{y}$ and $\mu_{\tilde{y}}^2$ in general form as follows:

$$\tilde{Y} = \overline{Y} + \sum_{\substack{i\neq j\neq k\neq \ldots \\ j<k<\ldots}} r_{iy.jk\ldots}\frac{\sigma_{y.jk\ldots}}{\sigma_{i.jk\ldots}}(X_i-\overline{X}_i) \tag{38}$$

$$\mu_{\tilde{y}}^2 = \frac{\sigma_{y.123\ldots s}^2}{n-s-1}\left[1 + \frac{\sum\limits_{\substack{i\neq j\neq k\neq l\neq\ldots \\ j<k<l<\ldots}}(X_i-\overline{X}_i)^2}{\sigma_{i.jkl\ldots}^2} - 2\sum_{\substack{i\neq j\neq k\neq l\neq\ldots \\ \text{for } r:\, i<j,\, k<l<\ldots \\ \text{for } \sigma:\, j<k<l<\ldots}} r_{ij.kl\ldots}\frac{(X_i-\overline{X}_i)(X_j-\overline{X}_j)}{\sigma_{i.jkl\ldots}\,\sigma_{j.ikl\ldots}}\right] \tag{39}$$

where i = 1, 2, . . . , s; j = 1, 2, . . . , s; k = 1, 2, ... , s; l = 1, 2, . . . , s; n denotes the number of individuals and s denotes the number of independent variables.

THE "EQUATING" OF NON-PARALLEL TESTS*

WILLIAM H. ANGOFF
Educational Testing Service
Princeton, New Jersey

THE PURPOSE of this paper is to explore the feasibility of establishing a unique conversion table for the purpose of translating scores on one test to scores on another test of different psychological function. In particular, this problem will be examined with reference to the Scholastic Aptitude Test of the College Board, referred to here as the SAT, and the American Council on Education Psychological Examination, referred to here as the ACE. The first portion of this paper will attempt to define the sources of error in such a table, and will draw some conclusions regarding their implications for a table of comparable scores. The second portion will present some empirical evidence regarding some difficulties in setting up such a table.

In discussing the implications of developing tables of comparable scores, it will be profitable to draw the distinction between parallel and non-parallel tests. In the conversion of scores from one form to a parallel form of a test, such as one form of the ACE to another, or one form of the SAT to another, there is simply the problem of transforming the system of units. In such a case, we can consider the problem as directly analogous to the problem of conversion from centimeters to inches, from Centigrade to Fahrenheit, from pounds to grams, etc. Since the two kinds of measures in each case involve identical functions, the system of conversion is unique; there is one and only one conversion equation. On the other hand, the question of converting scores from different tests is another matter. Here the problems include those that are involved in setting up a table of equivalent scores for tests of identical function, and in addition include the difficulties resulting from lack of equivalence of function of the two tests in question.

It is possible to outline briefly the following sources of random error in a table of equivalent scores for tests of identical function. One source of error lies in the unreliability of the measuring instruments themselves, and affects the stability of the statistics obtained in the collection of data. A second source of error lies in the choice of samples used to establish the conversion line. Different samples will yield lines that differ randomly from one another. A third source of error in equivalent scores lies in the design of the equating experiment and the method of treating the data. The various methods of equating are not equally reliable.

In the problem of equivalent scores that are derived from parallel forms of a test, there is by definition a unique "true" line. Errors resulting from the unreliability of tests, from the sampling of people, and from design and method will only be random errors. The problem of "comparable" scores, on the other hand, that is, of converting scores from different tests, is more complex. Here, the psychological functions measured are different, so that the problem is not simply one of transforming units. To use the physical analogy again: the problem essentially calls for a conversion from height to weight. However, the statement of this problem rests on an inappropriate premise, that there is a single conversion system—and this is obviously not true. While it is certainly true that one can predict or estimate by regression equations the most likely weight of an individual of a given height, or predict scores on the SAT from knowledge of scores on the ACE, it is clear that in neither case is the prediction to be considered a transformation of unit systems. Furthermore, just as in the case of regression lines, conversion lines which purport to translate scores across tests of different function will, in fact, differ systematically, that is, they will differ in a predictable fashion, depending on certain crucial considerations. Some of these considerations can be outlined as follows:

1. Different lines will result when the methodological definitions of comparability are different; and such definitions must in any instance be purely arbitrary. For example, one might ask what the most likely score would be on y, given a particular score on x. This would be one definition. Or one might equate standard-score deviates on the two tests, or equivalent percentiles. These would be other definitions. Or one might, for example, equate scores that yielded equal predicted scores on one or another criterion. In this instance, of course, the use of different prediction criteria would yield different equations. There is probably an unlimited number of other definitions, each yielding a different conversion line. In contrast, all methods or definitions of equivalent scores in the case of parallel forms lead to the same conversion equations, i.e., except for random error.

---

*This paper was presented at the 1954 meetings of the American Psychological Association.

2. Different lines will result when different populations form the basis for deriving the tables of comparable scores. The line of relationship between two tests of different function will be one for, say, male engineering students, and quite another for, say, female arts students. This fact is immediately apparent if one were to examine mean scores of boys and girls on the Verbal and Mathematical Sections of the SAT. Girls characteristically score slightly higher than boys on Verbal, but they score considerably lower than boys on Mathematical. The "comparability" line (or conversion line) between Verbal and Mathematical scores as derived from male data would then be much different from that derived from female data. Conversion lines for tests of different function must, in fact, be different for groups that show different types of profiles on the tests, since the conversion lines themselves are a way of examining profiles on two tests, and therefore they must be normative in character.

3. Differential selection will have a pronounced effect on conversion lines. Let us say, for example, that we collect data on the SAT and the ACE for a group of individuals in order to set up a conversion table to go from scores on one of these tests to scores on the other tests, and let us say further that these individuals have been explicitly selected on the SAT. The distribution of resultant SAT scores will, of course, be sharply curtailed. The distribution of ACE scores, on the other hand, will only be moderately altered. A conversion line derived over such data will be far different from a line derived over a group of unselected students. In fact, even if students used for data collection have been selected on an outside variable, say high school grades, their resultant distributions on the two scores in question will be different. More specifically, they will be unequally curtailed in proportion to the relative correlations of the two scores with the selection variable, with the result that the mean and variance of one score will be more drastically altered than the mean and variance of the other score. A line of relationship derived even from these data would have limited value. In contrast, scores taken from parallel tests will be equally affected by selection on an external variable, with the result that there will be no systematic effect on the conversion equation.

It should be emphasized here that the foregoing problems inherent in the development of a table of comparable scores do not replace or preclude the kinds of error-sources that were cited earlier in connection with equivalent scores that are taken from parallel forms. The problems of comparable scores exist in addition to the problems of equivalent scores.

Added to the systematic effects of (1) definition and method, (2) choice of population, and (3) selection, there are other problems, namely those of interpretation in considering a table of comparable scores. For example, there is a constant danger that the converted scores will be assumed to possess the reliability and the validity of the scale to which they are converted. While it should go without saying that ACE scores, however well they are converted to the SAT scale, are still ACE scores and retain all the basic characteristics of the ACE, many test users will consider that once converted, they should behave in all respects like SAT scores. Obviously this is unjustified. The reliability and validity of ACE scores that are converted to the scale of the SAT will be no different from the reliability and validity of the original unconverted ACE scores, and will not resemble SAT data in the slightest. Also, if regression methods are used for establishing the conversion line, then the predicted SAT scores will not have the same distribution as original SAT scores; the predicted SAT scores will have a regressed distribution. The pooling of these two kinds of distributions, taken from actual and predicted scores would therefore be inappropriate. And yet it is for the very purpose of pooling such data, or considering them in conjunction, that such conversion tables are ordinarily called for. And again, if regression methods are used, then the prediction of grades from predicted SAT scores will involve double regression while the prediction of grades from original SAT scores will involve only single regression. These two types of systems should not be merged, although the need that has been expressed for developing such a conversion is, essentially, to merge the systems or (which is equivalent in this context) to compare directly scores taken from different tests.

To point up the fact that the SAT and the ACE, although presumably both measures of general scholastic aptitude, are actually not equivalent forms, it will be of interest to examine the differences in item content and the correlations between the tests. The SAT is, for one thing, a three-hour power test; the ACE is a one-hour speeded test. The SAT contains verbal analogies, verbal opposites, verbal completions, and reading comprehension; the ACE contains verbal analogies, same-opposite items, and definitions. The SAT contains arithmetic reasoning items, and algebra and geometry items involving elementary principles; the ACE contains arithmetic, number series, and spatial items (figure analogies). At least from a cursory point of view, it would appear that the designs of these two tests are not entirely alike.

The correlations among the parts of the SAT and the ACE and their reliabilities are given in Table I of the reference material. They are average figures taken in part from the booklet

"College Board Scores", published by the College Entrance Examination Board, and also from validity and equating studies conducted by Educational Testing Service, and are estimated for unselected candidate groups. The reliabilities for the SAT are internal-consistency coefficients (KR-20); the reliabilities for the ACE are modified parallel-forms coefficients. The numbers in parentheses are estimated correlations between true scores. Admittedly, the figures are only rough estimates; they are given only for purposes of discussion. However, it is felt that these estimates are not substantially different from what would be found in a carefully controlled study.

It is observed in Table I that none of the estimated correlations between true scores is very close to unity, as would be the case if the tests were parallel. The highest true-score correlation, .90 for SAT-Verbal vs. ACE-Linguistic still shows almost one-fifth of the true-score variance unaccounted for. Consequently, we can be sure that the considerations normally applicable to the problem of equivalent scores are not applicable in the present situation, since these tests are not parallel.

It is suggested that considerable attention be given to the non-unique character of tables that purport to translate scores across tests of different function. While there are no data available that have been collected for the express purpose of demonstrating this non-uniqueness, some data are available and can be used as reference. For the most part they were collected for validity purposes, and do not quite fit the kind of design most desirable for demonstrating some of the points discussed above. However, within the limitations of these data, it will be seen that groups that would normally be expected to yield different profiles on the SAT and ACE do in fact yield different lines of comparability. It is primarily to the objective of multiple tables of comparable scores that the present data are addressed; and it is suggested here that as many tables need to be prepared as there are separate definable groups of individuals whose profiles on the two scores will be expected to be different, and who will normally be given the tests.

The present data were collected at five colleges: three liberal arts colleges referred to here as Colleges A, B, and C, a state university (College D) and a technological institution (College E). For the purposes of the present report the data are separated into two sets, those for the arts groups at Colleges A, B, and C, and those for four groups at the state university (College D). (The data taken from College E are based on ACE-Total score and will be considered later.) The separation into these two sets is occasioned by the fact that an unknown edition of the ACE was given at the state university (D). Consequently, those data cannot be shown simultaneously with the data from the other three schools.

Two groups of male liberal arts students are studied at College A, those entering college in 1949 (N = 398), and those entering college in 1950 (N = 448). Both groups took the 1949 form of the ACE College Edition. The form of the SAT probably varies within the groups, but its identification is of no consequence in these data, since all forms of the SAT are related to a common scale.

One group of 141 male liberal arts students is available from College B. This group entered college in 1948 and took the 1948 form of the ACE.

One group of 154 students is available from College C, and comprises male students in liberal arts and female students in a nursing school curriculum. Their year of college entrance was 1951; they took the 1948 form of the ACE.

Since conversion equations are available which permit the transformation of scores from the 1948 to the 1949 form of the ACE, it was possible to combine the foregoing four arts groups in one presentation.

Four groups of students are available at College D, the state university: 333 male liberal arts students, 582 female liberal arts students, 134 male students in Economics and Business, and 75 male engineering students. All students entered the university in 1947. The form of the test taken is unknown. The information available is that the ACE raw scores were converted to percentile scores and then transformed to a normalized scale with M = 13 and $\sigma = 4$.

A separate set of data are available for 305 freshmen entering the technological institution (College E) in 1950. For these students only SAT-M and ACE-Total (1948 Edition) scores are available. Consequently, these data are given separately from the rest.

Unfortunately, detailed knowledge is not available regarding the selection procedures at the five colleges. It is known that the ACE was taken after admission at College A; the SAT was taken before admission in the regular CEEB series. However, SAT scores were reportedly not used for selection for the 1949 entering group, and only to a "limited extent" for the 1950 entering group. At College B the SAT was given after entrance; nothing is known regarding the date of the ACE administration. At College C both tests were given after entrance. This appears to be similarly true at College D. Finally, at College E both tests were given after entrance; selection was made on the basis of high school grade.

It appears likely that in no case were the students selected explicitly on either the SAT or the ACE.

Seven tables are given, all derived simply by equating means and standard deviations. The first three tables describe the conversions obtained for the arts groups—two at College A, one

TABLE I

ESTIMATED CORRELATIONS, TRUE-SCORE CORRELATIONS, AND RELIABILITIES FOR SAT AND ACE

	SAT-V	SAT-M	ACE-L	ACE-Q	Reliabilities
SAT-V		.60 (.65)	.82 (.90)	.35 (.41)	.94
SAT-M			.42 (.48)	.63 (.71)	.90
ACE-L				.47 (.54)	.88
ACE-Q					.86

TABLE II

CONVERSIONS FOR COLLEGES A,B, AND C---ACE-L vs. SAT-V

ACE-L Scores	SAT-V Scores				
	College A (1949 Group) N = 398	College A (1950 Group) N = 448	College B N = 141	College C N = 154	Range of Converted Scores
40	236	239	247	253	17
60	376	376	385	395	19
80	516	512	522	538	26
100	656	649	659	681	32
120	796	786	796	824	38

TABLE III

CONVERSIONS FOR COLLEGES A, B, AND C---ACE-Q vs. SAT-M

ACE-Q Scores	SAT-M Scores				
	College A (1949 Group) N = 398	College A (1950 Group) N = 448	College B N = 141	College C N = 154	Range of Converted Scores
20	258	232	214	208	50
35	404	399	376	377	28
50	550	566	538	545	28
65	697	733	699	713	36
80	(843)	(900)	(870)	(882)	(57)

TABLE IV

CONVERSIONS FOR COLLEGES A, B, C,---ACE-Q vs. SAT-V

ACE-Q Scores	SAT-V Scores				
	College A (1949 Group)	College A (1950 Group)	College B	College C	Range of Converted Scores
	N = 398	N = 448	N = 141	N = 154	
20	(185)	225	(195)	(179)	46
35	355	379	355	346	33
50	526	532	516	512	20
65	697	686	676	679	21
80	(867)	(840)	(836)	(846)	(31)

TABLE V

CONVERSIONS FOR COLLEGE D---ACE-L vs. SAT-V

ACE-L Scores*	SAT-V Scores				
	Men Arts	Women Arts	Men E&B	Men Eng.	Range of Converted Scores
	N = 333	N = 582	N = 134	N = 75	
4	275	273	285	202	83
10	416	406	412	388	28
16	557	540	540	574	34
22	698	674	668	760	92
28	(839)	(807)	795	(946)	151

*Special scale used by College D.

TABLE VI

CONVERSIONS FOR COLLEGE D---ACE-Q vs. SAT-M

ACE-Q Scores*	SAT-M Scores				
	Men Arts	Women Arts	Men E&B	Men Eng.	Range of Converted Scores
	N = 333	N = 582	N = 134	N = 75	
4	(193)	(177)	246	281	104
10	359	323	382	429	106
16	525	469	519	577	108
22	691	615	656	725	110
28	(857)	761	792	(873)	112

*Special scale used by College D.

TABLE VII

CONVERSIONS FOR COLLEGE D---ACE-Q vs. SAT-V

ACE-Q Scores*	SAT-V Scores				
	Men Arts	Women Arts	Men E&B	Men Eng.	Range of Converted Scores
	N = 333	N = 582	N = 134	N = 75	
4	(149)	214	243	(109)	134
10	337	392	370	295	97
16	525	570	498	481	89
22	713	748	626	667	122
28	(901)	(926)	753	(853)	173

*Special scale used by College D.

TABLE VIII

CONVERSIONS FOR COLLEGE E (N - 305)
ACE-TOTAL vs. SAT-M

ACE-Total Scores	SAT-M Scores
20	215
60	357
100	502
140	645
180	790

at College B, and one at College C—separately for converting ACE-L to SAT-V, ACE-Q to SAT-M, and ACE-Q to SAT-V. (The last of these three is admittedly not one which would ordinarily be called for; however, it is considered useful in the present context in order to illustrate the effect of conversion with two tests that are known to have low correlation.) The second three tables parallel the first three except that they have been drawn up from data taken from College D for four different groups: Men in Arts, Women in Arts, Men in Engineering, and Men in Economics and Business. The last table describes the conversion from ACE-Total Scores to SAT-M, and is based on data collected at College E.

The three tables for the groups at Colleges A, B, and C describe a series of conversions which show a reasonable amount of agreement, but less agreement, it is judged, than would be found if the two tests were parallel. On the other hand, they are not so disparate that they would not be useful to colleges in providing rough score transformation.

It is observed in the foregoing three tables that the largest range of converted scores is 50 (in the ACE-Q vs. SAT-M conversion) and that the smallest is 17 (in the ACE-L vs. SAT-V conversion). In general, the ACE-L vs. SAT-V conversion shows the smallest amount of disagreement for the four groups of students, possibly because the correlation between those two scores is the highest of the three. In general, it is possible that disagreements of the magnitude observed here are not excessive, and could be tolerated. It is maintained that the small amount of disagreement among the lines is accounted for by the fact that the four groups in question are relatively homogeneous in character.

In contrast to the foregoing tables. Tables V, VI, and VII, show quite a different picture. These tables give similar types of information for the groups at College D, the state college. These groups are not considered to be homogeneous. As maintained above, they include liberal arts men, liberal arts women, men in economics and business, and men in engineering.

It is seen that the ranges of converted scores for the groups of students at College D are far greater than for the groups at Colleges A, B, and C. In spite of the fact that the groups at College D are all students in one college, their conversion lines differ more than the lines for the liberal arts students who come from different colleges. In respect to the type of conversion obtained, there is greater within-college variability than between-college variability. It is the judgment of the writer that this finding is wholly explainable by the differences in type of curriculum within the college. Liberal arts students show conversion lines that are characteristically different from those of engineering students, and male students show conversion lines that are characteristically different from those of female students. These differences occur when the tests in question are not parallel forms.

Finally, it will be of interest to examine the conversions obtained with data from College E, the technological institution. These conversions are summarized briefly in Table VIII.

It is striking to note in Table VIII that an ACE-Total score of 100 is found to be approximately comparable to 500 on SAT-M. If one were to generalize this conversion to all possible types of populations, it would be necessary to conclude that the norms group for the ACE, which normally averages 100 or so on the ACE, will score at 500 on SAT-M, and is therefore at the same general level of ability as the College Board standard group. However, in the light of the composition of the College Board college population and also of the composition of the usual ACE norms group, it must be admitted that the conclusion of similarity of general ability in the two populations is untenable. Consequently, the above conversion cannot be considered a general law operating for all possible types of student populations and must be considered specific for the group studied, or at least for similar groups.

Summary

The present paper has addressed itself primarily to the argument that conversion lines which relate tests of different function are primarily normative in character, and will not have generality of meaning. Different groups of individuals with their own characteristic backgrounds, curricular emphases, and interests will yield quite different types of conversion systems—and these conversion systems will reflect the differences in these groups. Consequently, it is considered erroneous to speak of "a conversion line" for tests of different function. Other errors of bias in addition to those resulting from differences in groups also accrue in this problem of non-parallel tests, such as those related to the effects of selection and those related to methodological definition—errors which do not appear in the problem of equivalent scores that are derived over parallel forms. Finally, errors of interpretation exist in the conversion of scores on non-parallel tests which are not present in the conversion of scores on parallel forms.

THE COMPARABILITY OF THE BI-FACTOR AND SECOND-ORDER FACTOR PATTERNS

JOHN SCHMID, Jr.*
Personnel Research Laboratory
Lackland Air Force Base, Texas

IN RECENT years a variety of tools for the multivariate analysis of psychological data have been developed (4). Among these, factor analysis has been used quite extensively. Unfortunately, however, factor analysis has been characterized by a multiplicity of forms, each yielding a different type of solution. The end of the proliferating process which has given rise to this assortment of solutions is not yet in sight, for new types of solutions are being developed yearly. The net result is that the research worker who would like to summarize a large group of intercorrelations in a meaningful way is confronted with apparently divergent methods and solutions.

It is time that we stop and take stock of some of the similarities and differences of various factor methods with the view that some unification is possible. It is the intent of this paper to demonstrate that under certain conditions two of the most popular solutions are identical except for computational methodology; that is, the bi-factor solution proposed by Holzinger (3) is identical to the second-order solution proposed by Thurstone (6).

In the bi-factor solution, the factors are orthogonal, consisting of one general factor—more appropriately designated as a basic factor by Burt (1: 46)—and group factors. The common factors of a nine variable bi-factor solution might appear symbolically as follows:

Variable	General Factor	Group Factors		
	b	I	II	III
1.	x	x		
2.	x	x		
3.	x	x		
4.	x		x	
5.	x		x	
6.	x		x	
7.	x			x
8.	x			x
9.	x			x

Another type of solution which has received widespread use is oblique simple structure. This has become popular because the patterning of the factor loadings gives the appearance of easy psychological interpretation. However, correlations between the factors offset this apparent gain in ease of interpretation. Thurstone, in recognizing this defect, offers an alternate solution in which the intercorrelations between the primary factors are factored completely. This complete solution, then, is used as a rotation matrix for transforming the oblique solution into an orthogonal solution in which the group factors of simple structure are maintained. However, additional factors are produced which are a derivation of the original oblique factor intercorrelations.

When one second-order factor exists, this solution resembles the stepladder pattern of the Holzinger bi-factor loadings. Thomson (5: Ch. XIX) illustrates the development of a complete second-order solution very clearly.

A survey of the three tests by Holzinger (3), Thurstone (6), and Thomson (5) has shown no explicit statement to the effect that the bi-factor solution and the second-order solution are identical, although statements tacitly suggest similarity of the two solutions. The purpose of this study, therefore, has been to gather evidence about the identity or similarity of the two solutions.

Holzinger has shown that the various factor solutions merely represent different sets of reference axes placed in the common factor space of the test variables, and he shows that transformation matrices exist which will carry one solution into another. He presents an illustration (3:257-258) whereby an oblique primary pattern, which is related very simply to a simple structure (2), is transformed into a bi-factor solution. He indicates (3:258) that a primary solution tends to have smaller communalities than its corresponding bi-factor solution because its common factor space is one dimension smaller. Question, however, arises if the number of dimensions in a common factor solution does affect the communalities. In the following example, some evidence is presented to show that, relative to the model used in this example, the communalities are unaffected by the added dimension found in a bi-factor solution.

In order to investigate the identity of the bi-factor and second-order solution, a model correlation matrix was constructed. It was believed that an empirical correlation matrix, definitely having only one second-order factor which would

---

*Air Force Personnel and Training Research Center, Air Research and Development Command

TABLE I

INTERCORRELATIONS OF VARIABLES*

$R = P\Phi P'$

	1	2	3	4	5	6	7	8	9	10	11	12
1.	0900	1800	2400	2700	0315	0441	0504	0567	0081	0243	0643	0729
2.		3600	4800	5400	0630	0882	1008	1134	0162	0486	1296	1458
3.			6400	7200	0840	1176	1344	1512	0216	0648	1728	1944
4.				8100	0945	1323	1512	1701	0243	0729	1944	2187
5.					2500	3500	4000	4500	0315	0945	2520	2835
6.						4900	5600	6300	0441	1323	3528	3969
7.							6400	7200	0504	1512	4032	4536
8.								8100	0567	1701	4536	5103
9.									0100	0300	0900	0900
10.										0900	2400	2700
11.											6400	7200
12.												8100

*Decimal points omitted

TABLE II

PRIMARY PATTERNS

Variable	P I	 II	 III
1.	.3		
2.	.6		
3.	.8		
4.	.9		
5.		.5	
6.		.7	
7.		.8	
8.		.9	
9.			.1
10.			.3
11.			.8
12.			.9

Intercorrelations of Factors

	Φ I	 II	 III
I	1.00	.21	.27
II	.21	1.00	.63
III	.27	.63	1.00

TABLE III

TRANSFORMATION MATRIX

g	E'		
	I	II	III
.3	.9539		
.7		.7141	
.9			.4359

TABLE IV

SECOND-ORDER FACTOR SOLUTION

Variable	b	B		
		I	II	III
1.	.09	.2862		
2.	.18	.5723		
3.	.24	.7631		
4.	.27	.8585		
5.	.35		.3570	
6.	.49		.4999	
7.	.56		.5713	
8.	.63		.6427	
9.	.09			.0460
10.	.27			.1308
11.	.72			.3487
12.	.81			.3923

be useful for this study, might be difficult to find in the literature. In addition, sampling errors of various types might tend to conceal a real identity under mere similar guise. Therefore, a model correlation matrix (Table I) was constructed by using the primary factor solution, P, and a one-common factor matrix of intercorrelations between the primary factor, ϕ (Table II); that is, the operation, $P\phi P'$, was performed to produce the correlation model. The intercorrelations between the factors, ϕ, were factored by the usual centroid method. One common and three unique factors were produced. These are reported as matrix E' (Table III). This matrix, E', is used as a transformation matrix for rotating the primary pattern, P, into a second-order solution. This appears as matrix B (Table IV). This second-order solution was computed in the usual way for constructing a second-order solution (6:Ch. XVIII; 5:Ch. XIX). At this point, we have a second-order factor solution which reproduces side correlations and communalities of our correlation matrix model exactly. The question to be answered in this study, now, was this: If our correlation matrix model, $P\phi P'$, is factored by the Holzinger bi-factor method, will a solution be obtained which is identical to the second-order factor solution which has just been constructed?

A Holzinger bi-factor solution was calculated independently on the correlation matrix model following the methodology prescribed by Holzinger (3). The factor loadings were found to be identical to those in Table IV. Thus, it was established, within the framework and limitations of the type of model set up in this problem, that there is no difference between the two types of solutions except for methodological procedure. This seems to reconciliate the apparent differences between the Holzinger and Thurstone school of thought on factor analysis. It should be noted that the communalities for the primary solution and the bi-factor solution are identical; so, apparently, communality is not a function of the number of dimensions used in obtaining either a simple structure solution or a bi-factor solution.

A second-order factor solution becomes considerably more complicated when more than one second-order factor is present. Speculation suggests that the factoring of higher order levels should be done successively with each resulting transformation, E_i, used to build a transformation matrix which will rotate each higher order oblique solution into bi-factor form.

REFERENCES

1. Burt, Cyril. "Group Factor Analysis," British Journal of Psychology, Statistical Section, III (1950), pp. 40-75.

2. Harris, C.W. and Knoell, D.M. "The Oblique Solution in Factor Analysis," Journal of Educational Psychology, (1948), pp. 385-403.

3. Holzinger, Karl J. Factor Analysis (Chicago: University of Chicago Press, 1951).

4. Rao, C. Radhakrishna. Advanced Statistical Methods in Biometric Research (New York: John Wiley and Sons, Inc., 1952).

5. Thomson, Godfrey H. The Factorial Analysis of Human Ability (New York: Houghton Mifflin Co., 1948).

6. Thurstone, L. L. Multiple-Factor Analysis (Chicago: University of Chicago Press, 1947).

AN ANALYSIS OF VARIANCE OF MULTIPLE MEASUREMENTS ON SUBJECTS CLASSIFIED IN UNEQUAL GROUPS OF ONE DIMENSION

RAYMOND O. COLLIER, Jr., and CLAYTON L. STUNKARD
University of Minnesota

Introduction

IN EXPERIMENTS where more than one measurement is taken on a subject, special analysis is required because of the lack of statistical independence among the measurements. Typical of such experiments are those in which subjects are first categorized into one of several different groups. Thereafter, multiple measurements are made on each subject under conditions corresponding to all combinations of one or more ways of classification.

Various writers have viewed aspects of the multiple measurements problem. The foundations were laid by Yates (8, 9, 10) in his early papers on the split-plot design which has been used extensively in agricultural experimentation. Among the more recent work on this general problem has been that by Moonan (5, 6), Hoyt, Stunkard, et al., (3), Kogan (4), Halperin (2), and Collier (1). This paper considers the model and analysis of a design in which subjects are classified into several groups with possibly unequal frequencies along one classification dimension. Then, each subject is measured under all combinations of a further two-way classification system.

The Model

Let us assume that N subjects are classified into R groups such that the r^{th} group contains n_r subjects with $r = 1, 2, \ldots, R$ and $\sum_r n_r = N$. For identification purposes this dimension, experimental or not, is designated classification dimension "r". Now, let the subjects within a group be designated by the subscript s with $s = 1, 2, \ldots, n_r$. Thus the subject is observed under all the combinations of P categories, represented by subscript $p(=1, 2, \ldots, P)$, of one classification dimension (dimension "p") with the Q categories, designated by $q(=1, 2, \ldots, Q)$, of a second way of classification which we name dimension "q". The random variable thus obtained is denoted by X_{pqrs} and refers to the measurement made on the s^{th} subject of the r^{th} group of dimension "r" on the p^{th} category of classification dimension "p" and the q^{th} category of classification dimension "q". Now, we assume that X_{pqrs} is PQN-variate, normally distributed with expected value a linear function of fixed constants,

$$E(X_{pqrs}) = \mu + \alpha(1,p) + \alpha(2,q) + \alpha(3,r) + \beta(1,pq) + \beta(2,pr) + \beta(3,qr) + \gamma(pqr) \qquad (1)$$

where μ is the general effect,

$\alpha(1,p)$, $\alpha(2,q)$, and $\alpha(3,r)$ are the main effects for the p^{th}, q^{th}, and r^{th} categories of dimensions "p", "q", and "r", respectively,

$\beta(1,pq)$, $\beta(2,pr)$, and $\beta(3,qr)$ are the interaction effects for the p^{th} and q^{th}, p^{th} and r^{th}, q^{th} and r^{th} categories of dimensions "p" and "q", "p" and "r", and "q" and "r", respectively;

$\gamma(pqr)$ are the interaction effects for the p^{th}, q^{th}, and r^{th} categories of dimensions "p", "q", and "r",

and again, as before

$p = 1, 2, \ldots, P$, $q = 1, 2, \ldots, Q$, $r = 1, 2, \ldots, R$, and $s = 1, 2, \ldots, n_r$.

The variance of X_{pqrs} is further assumed to be

$$V[X_{pqrs}] = \sigma^2 \qquad (2)$$

and any two measurements are assumed to be correlated with covariance

$$\mathrm{Cov}[X_{pqrs}, X_{p'q'r's'}] = \rho\sigma^2$$

where $\rho = 0$ when $r = r'$ and $s \neq s'$ or when $r \neq r'$ and $s = s'$.

The logic of this last assumption is that two or more measurements on the same subject are stochastically dependent, i.e., they have a covariance different from zero, whereas two measurements from different subjects are stochastically independent or have zero covariance.

Outline of Analysis

The analysis of the foregoing model requires that certain restrictions be placed on the parameters. Let these be

$$\sum_p \alpha(1,p) = \sum_q \alpha(2,q) = \sum_r n_r\alpha(3,r) = \sum_p \beta(1,pq)$$
$$= \sum_q \beta(1,pq) = \sum_p \beta(2,pr) = \sum_r n_r\beta(2,pr)$$
$$= \sum_q \beta(3,qr) = \sum_r n_r\beta(3,qr) = \sum_p \gamma(pqr) \qquad (4)$$
$$= \sum \gamma(pqr) = \sum_r n_r\gamma(pqr) = 0.$$

In order to render the dependent multiple measurements statistically independent of one another, a modification of a transformation due to Nandi (7) is applied to the X_{pqrs}. Letting a and b be particular values of p and q, respectively, we employ the following transformation:

$$Y_{abrs} = \sqrt{\frac{(a-1)Q+b}{(a-1)Q+b+1}}\left[X_{a(b+1)rs} - \frac{\sum_p^{(a-1)}\sum_q^{Q} X_{pqrs} + \sum_q^{b} X_{aqrs}}{(a-1)Q+b}\right]$$

where a = 1, 2, . . . , (P), and b = 1, 2, . . . , (Q-1),

$$Y_{abrs} = \sqrt{\frac{aQ}{aQ+1}}\left[X_{(a+1)1rs} \quad \frac{\sum_p^{a}\sum_q^{Q} X_{pqrs}}{aQ}\right]$$

where a = 1, 2, . . . , (P-1), and b = Q, (5)

$$Y_{PQrs} = \sqrt{PQ}\left[\frac{\sum_p^{P}\sum_q^{Q} X_{pqrs}}{PQ}\right]$$

As a result of this transformation it may be shown that the covariance of any two transformed variates is given by

$$\mathrm{Cov}[Y_{pqrs}, Y_{p'q'r's'}] = 0 \qquad (6)$$

In addition, the (PQ - 1)N transformed variates, Y_{abrs}, i.e., where p and q are not equal to P and Q simultaneously, have the variance

$$\sigma_1^2 = V[Y_{abrs}] = (1-\rho)\,\sigma^2 , \qquad (7)$$

and the N remaining variates, Y_{PQrs}, where p and q are equal to P and Q simultaneously, which may be labeled as Z_{PQrs}, have variance

$$\sigma_2^2 = V[Z_{PQrs}] = [1 + (PQ-1)\,\rho]\,\sigma^2 . \qquad (8)$$

The Y_{abrs} and the Z_{PQrs} then, are two variate sets normally and independently distributed within and between sets, but with different variances. Accordingly, we write two expressions for the sum of squares related to each of these sets, respectively:

(a) The Sum of Squares associated with Y_{abrs}

$$\sum_p\sum_q\sum_r\sum_s [Y_{pqrs} - E(Y_{pqrs})]^2 - \sum_r\sum_s [Y_{PQrs} - E(Y_{PQrs})]^2 = \qquad (9)$$
$$\sum_p\sum_q\sum_r\sum_s \Big\{X_{pqrs} - \overline{X}_{..rs} - [\alpha(1,p) + \alpha(2,q) + \beta(1,pq) + \beta(2,pr) + \beta(3,qr) + \gamma(pqr)]\Big\}^2$$

(b) The Sum of Squares associated with Z_{PQrs}

$$\sum_r\sum_s [Z_{PQrs} - E(Z_{PQrs})]^2 = PQ\sum_r\sum_s\Big\{\overline{X}_{..rs} - [\mu + \alpha(3,r)]\Big\}^2 . \qquad (10)$$

It should be pointed out here that the sum of squares in the right-hand side of (9) includes all the parameters of (1) except μ and $\alpha(3,r)$ while the right-hand side of (10) includes only these two parameters.

Employing normal regression theory or equivalently by the method of least squares quantities (9) and (10) above provide estimates of the parameters involved in each set. By well-known methods we may use these estimates in order to generate tests of hypotheses on the parameters. The quantities necessary to perform these tests are presented in usual analysis of variance form in Table I together with the related hypotheses.

First, we present tests of hypotheses on the parameters $\gamma(pqr)$, $\beta(1,pq)$, $\beta(2,pr)$, $\beta(3,qr)$, $\alpha(1,p)$ and $\alpha(2,q)$ involved in (9) above. The hypothesis concerning the interaction "p" × "q" × "r", $H_{(1)}$: $\gamma(pqr) = 0$, is tested by referring $F_{(1)} = \mathrm{M.S.}(\text{"p"}\times\text{"q"}\times\text{"r"})/\mathrm{M.S.}[\text{Error }(1)] =$

$$\frac{\mathrm{S.S.}(\text{"p"}\times\text{"q"}\times\text{"r"})/(P-1)(Q-1)(R-1)}{\mathrm{S.S.}[\text{Error }(1)]/(N-R)(PQ-1)}$$ to $F[(P-1)(Q-1)(R-1),\ (N-R)(PQ-1),\ \alpha]$, i.e., the upper

TABLE I

ANALYSIS OF VARIANCE FOR DOUBLE CLASSIFICATION OF MULTIPLE MEASUREMENTS ON SUBJECTS IN UNEQUAL GROUPS OF A SINGLE CLASSIFICATION[1]

Source of Variation	Hypothesis Tested	Degrees of Freedom	Sum of Squares[2]
Error (1)		(N-R)(PQ-1)	a - (b + f) + i
"p"x"q"x"r"	$H_{(1)}:\gamma(pqr)=0$	(P-1)(Q-1)(R-1)	b - (c + d + e) + (g + h + i) - j
"p"x"q"	$H_{(2)}:\beta(1,pq)=0$	(P-1)(Q-1)	c - (g + h) + j
"p"x"r"	$H_{(3)}:\beta(2,pr)=0$	(P-1)(R-1)	d - (g + i) + j
"q"x"r"	$H_{(4)}:\beta(3,qr)=0$	(Q-1)(R-1)	e - (h + i) + j
"p"	$H_{(5)}:\alpha(1,p)=0$	P-1	g - j
"q"	$H_{(6)}:\alpha(2,q)=0$	Q-1	h - j
Within Subjects		(PQ-1)N	a - f
Error (2)		N-R	f - i
"r"	$H_{(7)}:\alpha(3,r)=0$	R-1	i - j
Between Subjects		N-1	f - j
Total		PQN-1	a - j

[1] No Mean Square column is presented since the hypothesis tests are given specifically elsewhere. However, the expected values for the mean squares of Error (1) and Error (2) are found to be σ_1^2 and σ_2^2, respectively, as expressed in [7] and [8].

[2] The lettered quantities are defined as follows: $a = \sum_p \sum_q \sum_r \sum_s X_{pqrs}^2$,

$b = \sum_p \sum_q \sum_r [(\sum_s X_{pqrs})^2/n_r]$, $c = \sum_p \sum_q (\sum_r \sum_s X_{pqrs})^2/N$, $d = \sum_p \sum_r [(\sum_q \sum_s X_{pqrs})^2/Qn_r]$,

$e = \sum_q \sum_r [(\sum_p \sum_s X_{pqrs})^2/Pn_r]$, $f = \sum_r \sum_s (\sum_p \sum_q X_{pqrs})^2/PQ$, $g = \sum_p (\sum_q \sum_r \sum_s X_{pqrs})^2/QN$,

$h = \sum_q (\sum_p \sum_r \sum_s X_{pqrs})^2/PN$, $i = \sum_r [(\sum_p \sum_q \sum_s X_{pqrs})^2/PQn\ ]$, $j = (\sum_p \sum_q \sum_r \sum_s X_{pqrs})^2/PQN$

with $p=1,2,\cdots,P$, $q=1,2,\cdots,Q$, $r=1,2,\cdots,R$, $s=1,2,\cdots,n_r$ and $N = \sum_r n_r$.

TABLE II

RAW DATA FROM A RELEARNING EXPERIMENT

Group																				
Experimental (r=1)														Control (r=2)						
	List (p)							List (p)							List (p)					
	1			2				1			2				1			2		
Subject Number	Trial (q)			Trial (q)			Subject Number	Trial (q)			Trial (q)			Subject Number	Trial (q)			Trial (q)		
	1	2	3	1	2	3		1	2	3	1	2	3		1	2	3	1	2	3
s= 1	2	7	9	2	8	9	s=19	0	1	4	0	4	5	s= 1	5	6	10	4	7	10
2	5	8	10	7	10	10	20	1	6	9	2	8	7	2	4	7	9	4	5	10
3	1	6	8	1	6	9	21	2	2	4	2	3	5	3	4	5	9	5	5	10
4	4	6	6	2	4	8	22	0	7	10	5	8	10	4	3	10	10	6	10	10
5	1	5	10	4	8	10	23	4	6	7	2	5	8	5	3	3	6	2	5	6
6	2	5	7	2	2	7	24	2	5	4	4	4	6	6	1	3	2	2	4	5
7	3	6	8	3	6	9	25	1	2	4	1	2	6	7	4	6	9	5	8	8
8	1	5	9	1	4	8	26	2	5	9	3	7	8	8	6	10	10	4	10	10
9	4	4	5	3	7	8	27	2	3	6	0	3	5	9	4	5	10	1	4	9
10	6	10	10	6	8	10	28	3	8	9	5	5	9	10	1	2	5	1	0	4
11	2	2	4	2	4	5	29	1	6	9	2	6	8	11	7	10	10	8	9	10
12	1	2	6	1	4	5	30	1	4	6	2	3	6	12	3	8	10	3	7	9
13	1	2	7	2	5	7	31	3	9	10	4	9	10	13	3	3	6	4	6	7
14	2	7	9	5	7	9	32	4	5	8	2	7	8	14	2	2	7	4	8	8
15	0	5	9	5	8	9	33	4	8	10	3	9	10	15	2	6	10	4	9	9
16	1	6	8	3	5	4	34	2	4	6	4	8	9	16	2	5	8	2	6	8
17	4	7	10	1	7	10								17	1	3	4	4	4	4
18	3	4	10	3	8	9								18	2	5	9	2	6	8

TABLE III

MEANS BY LIST, TRIAL, AND GROUP FROM A RELEARNING EXPERIMENT

Group	n_r	List (p)	Trial (q)			
			1	2	3	All
Experimental (r=1)	34	1	2.21	5.24	7.65	5.03
		2	2.76	5.94	7.82	5.51
		Both	2.49	5.59	7.74	5.27
Control (r=2)	18	1	3.17	5.50	8.00	5.56
		2	3.61	6.28	8.06	5.98
		Both	3.39	5.89	8.03	5.77
Total	N=52	1	2.54	5.33	7.77	5.21
		2	3.06	6.06	7.90	5.67
		Both	2.80	5.69	7.84	5.44

TABLE IV

COMPLETE ANALYSIS OF VARIANCE OF DATA FROM A RELEARNING EXPERIMENT

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F	Test of Hypothesis[1]
Error (1)	250	411.1645	1.6447	-------	--------
List × Trial × Group ("p" × "q" × "r")	2	0.1409	0.0704	....	Accepted
Residual (1)	252	411.3054	1.6322	-------	--------
List × Trial ("p" × "q")	2	4.7501	2.3750	1.455	Accepted
List × Group ("p" × "r")	1	0.0525	0.0525	....	Accepted
Trial × Group ("q" × "r")	2	5.7832	2.8916	1.772	Accepted
List ("p")	1	16.6153	16.6153	10.180	Rejected
Trial ("q")	2	1,329.8269	664.9134	407.373	Rejected
Within Subjects	260	1,768.3334			
Error (2)	50	837.0512	16.7410	-------	--------
Group ("r")	1	17.5769	17.5769	1.050	Accepted
Between Subjects	51	854.6281			
Total	311	2,622.9615			

---

[1] A significance level of .01 was employed.

α-point of the F distribution with (P-1)(Q-1)(R-1) and (N-R)(PQ-1) degrees of freedom, since $F_{(1)}$ is, under H(1), distributed as F[(P-1)(Q-1)(R-1), (N-R)(PQ-1)].

If the γ(pqr) are zero, the hypothesis on the interaction "p" × "q", $H_{(2)}$: β(1, pq) = 0, is tested by referring $F_{(2)}$ = M. S. ("p" × "q")/M. S. [Residual (1)] =

$$\frac{\text{S. S. ("p"} \times \text{"q")}/(P-1)(Q-1)}{\{\text{S. S. [Error (1)]} + \text{S. S. ("p"} \times \text{"q"} \times \text{"r")}\}/[(P-1)(Q-1)(R-1) + (N-R)(PQ-1)]}$$

to F{(P-1)(Q-1), d. f. [Residual (1)] = (P-1)(Q-1)(R-1) + (N-R)(PQ-1)], α}. The two other hypotheses of the interaction "p" × "r", H(3): β(2, pr) = 0, and the interaction "q" × "r", $H_{(4)}$: β(3, qr) = 0, are tested in analogous form as H(2), the numerator only differing.

Concerning the main effect "p", the hypothesis, H(5): α(1, p) = 0, is tested by referring F(5) = M. S. ("p")/M. S. [Residual (1)] to F{(P-1), d. f. [Residual (1)], α}. Finally, the remaining hypothesis relative to the main effect "q", H(6): α(2, q) = 0 is tested in a parallel fashion as H(5).

Testing hypotheses on parameters μ and α(3, r) involved in (10), we consider first the hypothesis concerning the main effect "r", H(7): α(3, r) = 0. This test consists of referring F(7)=M. S. ("r")/MS. [Error (2)] to F[(R-1), (N-R), α]. Although the hypothesis on the general effect, $H_{(8)}$: μ = 0, has not been listed specifically in Table I, this hypothesis may be tested by referring $F_{(8)}$ = j/M. S. [Error (2)] to F[1, (N-R), α].

Special attention is called to the fact that hypotheses $H_{(1)}$ to $H_{(6)}$ representing the four interaction effects and the two main effects are to be tested with a different estimate of error than is the main effect related to classification dimension "r" used to group the subjects.

An Application of the Analysis

For an example of a specific problem in which the above type of analysis is appropriate we refer to a re-learning experiment by Carlson* in which he randomly assigned 52 female residents of a college dormitory to two groups, experimental (n_1 = 34), and control (n_2 = 18). Using two lists of nonsense words, Carlson provided the experimental group with special training during sleep while the control subjects were not so treated. This special training consisted of reading the ten words in List 1 with "correct" associates and the ten words in List 2 with certain "incorrect" associates. After his training period, the twenty words with the "correct" associates were read to both groups followed by an immediate testing in multiple choice form. This so-called re-learning procedure was continued for three trials. The measurements recorded for each subject consisted of the number of "correct" associations made during the testing on each list and trial.

Before applying the analysis presented in Table I, it is first necessary to recognize that the following assumptions are made. The variance of any observation is assumed to be σ^2, and any two observations taken on the same subject are assumed to be dependent with covariance equal to $\rho\sigma^2$. Any two observations on different subjects are assumed to be independent and hence have zero covariance.

The measurement taken on the s^{th} member of the r^{th} group of subjects on the q^{th} trial of the p^{th} list of words may be designated as Ypqrs, where p = 1, 2, q = 1, 2, 3, r = 1, 2, and s = 1, 2, . . . , n_r with n_1 = 34 and n_2 = 18. The raw data for this experiment are presented in Table II, while the means of the measurements are to be found in Table III. The basic values needed for the sums of squares in the analysis of variance outlined in Table I were computed and found to be as shown on the next page. Substitution of these values in the expressions of Table I together with the proper values (N = 52, P = 2, Q = 3, and R = 2) for computing the degrees of freedom, and proceeding to test the null hypotheses as outlined permits presentation of the complete analysis of variance in Table IV.

Examination of the table of means (Table III) together with the results of the tests of the various hypotheses (Table IV) leads to the conclusion that List 2 is more difficult to learn than List 1, and that significant gains accrue from trial to trial. However, the experimental group does not perform significantly better or poorer than the control group, and none of the three ways of classifying the measurements interacts with the others, that is, the classifications of List ("p"), Trial ("q"), and Group ("r") are independent of each other.

Summary

A model is set forth for a specific class of experimental designs in which subjects classified into several groups with possibly unequal frequencies along one dimension are measured under all combinations of categories of a further two-way classification system. Because the multiple measurements on each subject are logically assumed to be dependent, they are trans-

---

*Raymond Carlson, Dean of Students, Bemidji State Teachers College, Bemidji, Minnesota.

a = 11,864.0000
b = 10,615.7843
c = 10,592.2308
d = 9,275.2832
e = 10,594.2255
f = 10,095.6666
g = 9,257.6538
h = 10,570.8654
i = 9,258.6154
j = 9,241.0385

formed into independent variates. The transformation, an outline of the analysis, and the appropriate analysis of variance with the necessary equational quantities to use in testing hypotheses for the general case are presented. To illustrate the procedure, data and analysis for a special case of the design are also presented.

REFERENCES

1. Collier, R. O., Jr. *Experimental Designs in Which the Observations Are Assumed to Be Correlated*, unpublished Ph. D. Dissertation, University of Minnesota, 1956.
2. Halperin, Max. "Normal Regression Theory in the Presence of Intra-Class Correlation," *Annals of Mathematical Statistics*, XXII (1951), pp. 573-80.
3. Hoyt, C. J., Stunkard, C. L., *et al*. "A Comparison of Two Methods of Instruction in Beginning Drawing," *Journal of Experimental Education*, XX (1952), pp. 265-79.
4. Kogan, L. S. "Analysis of Variance—Repeated Measurements," *Psychological Bulletin*, XLV (1948), pp. 131-43.
5. Moonan, W. J. "Simultaneous Examination and Method Analysis by Variance Algebra," *Journal of Experimental Education*, XXIII (1955), pp. 253-57.
6. Moonan, W. J. "An Analysis of Variance Method for Determining the External and Internal Consistency of an Examination," *Journal of Experimental Education*, XXIV (1956), pp. 239-44.
7. Nandi, H. K. "A Mathematical Set-up Leading to Analysis of a Class of Designs," *Sankhyā*, VIII (1947), pp. 172-76.
8. Yates, F. "The Principles of Orthogonality and Confounding in Replicated Experiments," *Journal of Agricultural Science*, XXIII (1933), pp. 109-45.
9. Yates, F. "Complex Experiments," *Supplement to the Journal of the Royal Statistical Society*, II (1935), pp. 181-223.
10. Yates, F. *The Design and Analysis of Factorial Experiments*, Technical Communication No. 35 (Harpenden, England: Imperial Bureau of Soil Science, 1937).

Three other articles that were scheduled for the March issue of the *Journal of Experimental Education* will be published in the March issue

Journal of Experimental Education

Volume XXV June, 1957 Number 4

A PROGRAM FOR THE SELECTION OF TRAINEES IN PHYSICAL MEDICINE

NICK JOHN TOPETZES*
Marquette University
Milwaukee, Wisconsin

INTRODUCTION

SINCE the end of World War II many perplexing problems concerning the need and purpose of education in our modern, complex society have been brought to public attention. The whole aim of education in a democracy is to equip each individual student for a better life and to guide him so he may contribute to the betterment of society. In recent years, educators have come more and more to respect the importance of individual differences. No two students are exactly alike in what they need to get through education to enable them to live fully. To meet these varying needs is not an easy task for educational institutions, especially since the students, themselves, are seldom aware of their own particular needs, and therefore depend largely on the advice of parents, teachers and others. Often the products of our American schools and colleges have reached adulthood, unhappy in the work in which they find themselves, and discontented, therefore, with their lot in life. Surely for them education has failed to fulfill its aim: to equip the individual for better living.

Realizing that such conditions have prevailed and that American education has not done all within its power to produce well adjusted individuals, more educators in recent years have put forth special effort by evaluating their own work and experimenting with new ideas and procedures. Out of their evaluation and experimentation has come, among other things, an organized program of student guidance, based, to some extent, on objective data from psychological measures, intelligence, achievement, judgment, interest, personality adjustment, aptitude, and other tests. Such a testing program, combined with interpretation of the results and information about appropriate vocational fields, has led many people into work they might not otherwise have done, but for which they are found to have the necessary qualifications. The result is that their whole educational program focuses on a definite aim, and in turn, they find the satisfaction that only those who are doing the work for which they are best suited can know.

Personnel work in a college or university is the systematic bringing to bear on the individual student all those influences, of whatever nature, which will stimulate him and assist him, through his own efforts, to develop in body, mind and character to the limit of his individual capacity, ability, interests, and aptitudes for growth.''(108)

Research must be brought to bear upon these problems of preparing, assisting, and stimulating each student in selecting the occupation for which he or she is best fitted, in the light of his or her total educational, vocational and social potentialities.

> ''It is costly for society to spend time and effort in educating individuals to do tasks which they, at best, will perform on a very low level of efficiency. Both the individual and society have a stake in education.'' (87)

The great growth in college enrollments, the changes in the availability of positions for college graduates, and the development in the science of

---

*The writer is especially indebted to Professor A. H. Edgerton, University of Wisconsin, for his invaluable and constant encouragement throughout the course of this investigation, also sincere thanks are due Dr. Elizabeth Grimm, Dr. Harry D. Bouman, Miss Margaret A. Kohli, and Miss Caroline G. Thompson of the University of Wisconsin who contributed pertinent information to the furtherance of this project. Appreciation is extended to the National Foundation for Infantile Paralysis whose financial assistance made this research study possible. Much help was received from the faculty members of the Department of Education of the University of Wisconsin, particularly from Professors Gregg, Rothney, and Gerth. Mention should be made of those former graduate students at the University who gave so freely of their criticism and suggestions during the progress of this study.

Bureau Edl. Research
Dated ....
Acc. No ....

testing go hand-in-hand in making the problem of prediction of vocational success of greater importance than ever before. The need for the improvement of methods of prediction and selection for the guidance of entering college students is evident in nearly every phase of modern professional life. Higher education must become increasingly efficient by basing its plans and practices on the positive objective findings of statistical, laboratory, and experimental studies.

This investigation points the way for increasing the accuracy of knowledge about the possibilities of achievement and success of some of our entering college freshmen who desire to enter physical medicine.

Consider, for instance, the students who intend to become physical or occupational therapists. There is room and need in society for men and women who are able practitioners of physical medicine. There is hardship, however, on the one hand, for the student trainee who finally gets through and is ill-fitted to the profession, and, on the other hand, for the disabled public when served by poor practitioners.

Harold Clark and others (15) have studied occupational distribution and trends. They have found that there is no close relationship between the number of persons preparing for an occupation and the number of persons needed each year by the occupation in question. There consequently is seldom a close relationship between the supply of workers prepared for an occupation and the number of qualified employees the occupation can actually absorb.

This fact is important because it shows clearly that our prevailing methods of vocational guidance ordinarily are trustworthy predictions of success or accurate data which are not used sufficiently on labor markets and employment objectives. Since it is society, as well as the individual, who will benefit, it is an obligation of the schools and colleges to make objective and subjective measures of the student's abilities, capacities, interests, personality and social adjustment, and aptitudes. Assume that there are 200 individuals wishing to prepare for a certain vocation; suppose further that through the application of aptitude and related measures and other available data the department's faculty finds that only 100 of the individuals can profitably prepare themselves for the vocation. Some error is certain to result in this selection, so that of the 100 selected not all will prove to be of the most promising 100. Some of the rejected 100 will be those who would have made good in the preparation and training for the vocation and in the vocation itself.

There are a few fields in which guidance can be conducted with considerable promise of success. These fields are recognized as professions, which require some lengthy preparation. This fact alone operates as a method of selection, since only those who can afford the time and money involved in such preparation can be expected to enter it. From the point of view both of the individual student and of service to society, something more than such a chance or haphazard method of selection is needed. The individual students should be protected against unnecessary disappointment and disillusionment which result from failure after once being admitted to an institution of higher learning for professional preparation. The interests of society should be safeguarded against inefficient persons by the maintenance of adequate educational, vocational, and social standards. It is the duty of our higher institutions to discriminate adequately between those who show high promise and those who offer little likelihood of success.

SECTION I

NATURE AND SCOPE OF THE STUDY

A. Nature of the Problem

FOR EFFECTIVE guidance of students into college work, articulation between the student's capacities, interests, aptitudes, personality, and possibilities of achievement is necessary. The place of the college in the educational scheme is an important item in considering the problem of selection, admission and prediction. Guidance of pupils from high school into college should be as thorough and complete as possible. College to most people is a preparation directly for vocations which requires serious consideration from both student and faculty.

> "No subjective or objective method of estimating success has as yet been inaugurated which is perfect. It is true that the most valid method for finding out whether or not a student can achieve a certain scholastic standing or be successful in an occupation is for her to try it. If, for instance, a student wants to become an engineer, the most valid method for him to find out if he can become one is first to attempt the training necessary and then go out and attempt the job. This trial and error method, however, is costly both for the individual and the school, let alone society. For the individual it is a costly procedure because he may spend much time in trying out various lines of work." (24)

The general problem of this investigation is the evaluation of certain factors for estimating the future performance of individuals applying for admission to or having actually entered the physical or occupational therapy courses at the University of Wisconsin. In other words, the problem is to determine how to use best the knowledge of a number of characteristics and data of students entering the Department of Physical Medicine in predicting their future success as measured by certain selected criteria. The problem was investigated to determine more accurate means of predicting success and to set up methods for interpreting the data resulting which might be used by the faculty members in Physical Medicine for selection and guidance of their students.

B. Purpose of the Study

To make a wise selection from the list of applicants to occupational and physical therapy courses has long been recognized as a formidable task. Because of considerable dissatisfaction with the results of previous methods of selection, the Department of Physical Medicine has been conducting this cooperative study with the use of several tests (intelligence, interest, achievement, special aptitudes, personality and social adjustment, and judgment), a rating scale, and a data form to supplement other methods of selection of students.

The purpose of this study is to discover sufficiently reliable, and valid means for predicting the success of students entering Physical Medicine at the University of Wisconsin, knowing that these predictors will help in the initial and continuous selection of students. If the selection of students for physical medicine can be more efficiently made, much waste for the institution and disappointment for the individual can be prevented. Selective devices are necessary for Physical Medicine.

The problem in the initial selection of students is to develop a testing program which will provide valid and reliable information (objective and subjective data) about the intelligence, achievement, interest, special aptitudes, personality and social adjustment, and reasoning and judgment of each student in physical therapy and occupational therapy.

The general purpose of the testing program is to provide data which will help in the solution of the problem of selecting trainees in occupational and physical therapy. The purpose of the study through the use of the tests, a rating scale, grade point average, and predicted grade point average has established the following objectives:

1. To provide accurate and meaningful measurements of the individual's intelligence, interest, achievement, special aptitudes, personality traits, emotional adjustment, and judgment.
2. To assist in the selection of students with potentialities to make good physical or occupational therapists, and to arrange guidance towards other fields for those who seem unsuitable for such specialization.
3. To use measuring devices which will reduce the amount of failures and consequent elimination of students from the training courses, since the latter, of necessity, have had to establish enrollment limits.
4. To provide diagnostic and remedial measures which will keep maladjustment at a minimum.
5. To develop a practical means of predicting success in physical or occupational therapy from combining the student's scores on a test battery and other pertinent data.
6. To install a basic functional guidance and testing program which can be developed as an integral and functioning part of the total school program of every physical and occupational therapist.
7. To select and/or devise professional and standardized tests which, in combination with each other or with other measures, may form a selective measure and prediction of the efficiency of therapists in Physical Medicine.

The present study has been made possible by the cooperation of the University of Wisconsin Medical School, the Department of Education, and the National Foundation for Infantile Paralysis. * This investigation was begun during the second semester of the 1945-46 school year and continued through both semesters of the 1946-47 school year at the University of Wisconsin.

The professional divisions of medicine are interested in having the various departments of Physical Medicine, occupational, and physical therapy, supply competent and well-trained therapists. In order to do this, the schools and their faculties understand that they must attract and select for training those students who are most likely to succeed and become able therapists. All of the schools of physical medicine employ some method for selecting their students from the applicants who seek admission. Most of them study the student's previous school record, many insist upon a personal interview, and some demand that the student demonstrate a certain competence in specific examinations.

Some schools of occupational and physical therapy have not had as many applicants as they would wish for and this may have caused a few to accept students whose chances for becoming competent therapists were not very promising.

*The National Foundation For Infantile Paralysis provided all of the financial assistance needed for this research project.

While this has undoubtedly acted as a challenge to the schools and has made them redouble their efforts to produce the best possible therapists from these poorer students, yet Physical Medicine desires that the efforts of its schools be concentrated on training those who are most likely to become top-quality therapists.

It has been the hope of many individuals, groups, and committees during the past years that attention be given to the use of aptitude tests for improving the selection of therapy students. In 1945, the Medical School and the Department of Education of the University of Wisconsin appointed a committee to study this problem.

In February, 1946, a committee comprised of Dr. Alanson H. Edgerton, Professor of Education, who is in charge of Guidance Training courses; Dr. Elizabeth Grimm, Medical Director of the Department of Physical Medicine; Dr. Nick J. Topetzes, assistant professor of Education, Marquette University, and author of this study; Miss Caroline G. Thompson, Technical Director of the School of Occupational Therapy; and Miss Margaret A. Kohli, Technical Director of the School of Physical Therapy, decided after considerable discussion to use twelve tests covering the areas of intelligence, interest, achievement, special aptitudes, personality traits and social adjustment, and practical judgments.

It was felt that the place to start was with the administration of this battery of tests, already accepted as reliable and significant, from which a profile on each student could be drawn. The results of these tests were to be compared with the subjective rating of the students by their instructors and clinical supervisors during their training period and with the rating given them by their employers and supervisors at the end of their first year of vocational experience. The rating form which was used is discussed in Section III. It was hoped by such correlation that certain tests would stand out as being of significance in the evaluation of the aptitude of the student. If these tests would not accomplish this purpose it would then be necessary to develop one battery which will.

In the selection of students for Physical Medicine (especially with the current influx) the University is concerned with maintaining high standards for admitting students into the Department. That is why the need of some methods of selective admission to the study of physical and occupational therapy has arisen. The increasing number of students in Physical Medicine, the limitation of clinical facilities and the desire to prevent overcrowding in the schools made it seem necessary that an experimental guidance and testing program be developed for this Department. The Department of Physical Medicine naturally desires the student who is likely to succeed in the field because he or she has the potentialities to make a good therapist and the safest guarantee is the incorporation into the Department of a basic testing and guidance program with selective devices which possess the greatest predictive value.

C. Review of Related Literature

Many studies have been made related to the selection and prediction of students in college. This problem of college success and vocational success has been confronting the schools in our democratic society for many years. Educators have been probing into this matter hoping to find some reliable and valid means for selecting students and predicting college and vocational success.

Douglass and associates (25) of the University of Minnesota found that:

1. Success can be predicted more accurately in curriculums to which first-year college students are admitted, e.g., engineering, nursing, pharmacy, agriculture, and in the School of Business Administration and College of Dentistry (coefficients of multiple correlation run from .70 to .80), and less accurately in the other curriculums to which students are admitted only after two or more years of college work, e.g., law, medicine, business administration, and education (best coefficiencies of multiple correlation run from .60 to .70).
2. Prediction for vocational success can be made more accurately for the poorest and the ablest students than for those of medium ability.
3. Tests of general mental ability and achievement not only are less closely correlated with scholastic success than the more pertinent special aptitude tests, but they add definitely less to multiple correlation coefficients.
4. Interest scores were found to be only slightly correlated with marks and to add nothing to other variables for prediction.
5. Two general reading tests (speed) were found to be less closely correlated with success than other variables requiring less time.

Segel (93) found that the correlations between general mental tests and achievement tests are lower than those between the same tests and general college scholarship. Those with general achievement tests are likewise slightly lower while those with specific traits, interests, aptitudes, or achievements are approximately the same as the relationships between mental tests and grade point average. These data led Segel to the conclusion that for predicting success in a college program, tests of specific aptitudes, interests, and achievements are the best.

Although the number of individuals engaged in various professions is small compared to the total number of persons employed, these occupations represent the vocational goal of many who cannot possibly succeed. The fact that the occupations and professions require college training has acted as a selective factor which barred many who were incapable of succeeding. Fortunately, the recent expansion of higher education intensifies the need for accurate predictions of those most likely to be successful in professional training courses and in subsequent occupational adjustment.

A definite level of scholarship, results of general intelligence tests, and personality characteristics have been general criteria used to indicate success in professional courses like Medicine, Nursing, Law, Dentistry, and Business Administration. In addition to these general criteria instruments of prognosis, rating scales, special aptitude tests, and judgment scales have been developed to indicate specific professional aptitudes.

Because medical schools have always been selective, rigorous standards of scholarship accompanied by personal characteristics as judged by members of the schools have been used as criteria of selection. Recently, however, the Association of American Medical Colleges has been using a prognostic test devised by Moss (71) which consists of comprehension and retention, visual memory, memory for content, logical reasoning, and scientific vocabulary. It is now being given in 600 colleges. The scores bear a significant relationship to success in medical school. The Medical Aptitude Test scores predicted 53 percent of the failures and excluded only 5 percent of those with high scholarship records. A combination of the test with premedical scholarship data predicted 69 percent of failures and excluded only 4 percent of those with high average marks.

Because dentistry seems to require more mechanical and motor ability than medicine, tests of mechanical aptitude and manual desterity have been used to predict success in this field. In this field, Schultz (92) gave tests of motor adaptability, motor learning, and general intelligence to second-year dental students. The motor-adaptability score yielded the highest validity coefficients (.25 to .37). Harris (43), however, reported that five mechanical-aptitude tests gave only chance correlations with his two criteria of success in dental school. The Otis Intelligence Test correlated .55 with first year average marks and, when combined with predental scholarship, gave a multiple correlation of .67 with average dental-school marks. It is still problematical whether motor and mechanical ability tests increase the accuracy of prediction of success in dental schools.

Douglass and McCullough (25) made a study of 83 students at the University of Minnesota showing the relation between the predicted dental honor point ratio and the actual first-year dental grades. These predictions were made on the basis of predental grades, and scores on mental ability test, metal filing test, visual memory test, and finger dexterity tests.

MacPhail and Joslin (65) found that 124 girls in training at the Rhode Island Hospital were superior to 124 students in training at another institution because the students of nursing education at Rhode Island Hospital were selected by means of intelligence, reasoning, personality, and achievement tests.

In Law, the Ferson and Stoddard Law Aptitude Examination is given, although similar tests have been developed at Yale, Columbia, and other universities. It measures capacity for accurate recall, comprehension of passages read, ability to reason by analogy, ability to reason analytically, and skill in pure logic. Research has shown that students ranking below the median on the test subsequently fail in such large numbers that such scores may be interpreted as indicating lack of legal aptitude.

Gaudet and Marryatt (35) found that this test was better than freshman marks in predicting marks in law school.

Crawford (19) reports that the School of Law of Yale University adopted the practice of selective admissions, and about 1926 limited the number to be admitted each year to 100. In addition to the general requirements—college record, letters of recommendation from college instructors, and interviews—experiments were undertaken with aptitude tests. The first tests used were the Thorndike CAVD Test, which took four hours and a test of twenty minutes constructed by Chapman. The average correlation of the Thorndike Test was +.45, of the Chapman Test +.40, both lower than the correlation between college grades and first year grades in law school. Neither test, however, was a test of legal aptitude specifically.

In 1937 the Law School of the University of California adopted the use of a legal aptitude test for purposes of selection and as a possible predictor of success. Dickinson (23) found after a period of four years' experimentation that students who scored high on the aptitude test did well in the Law School's course program. The correlation of the aptitude test with grades in the Law School was .75.

The development and application of tests for selection of students to study engineering have had a longer history than either medical or law aptitude tests.

An experiment was undertaken by Hammond and Stoddard (42) to discover whether aptitude tests were useful as means of predicting probability of success in engineering. The correlation between composite scores in the entire series of English, Physics, Chemistry, and Mathematics Aptitude Tests and general success in the work of the freshman year was .70.

Sackett (90) reported that aptitude tests in English and Mathematics were given to nearly 3000 entering students in twelve colleges. The mathematics aptitude test showed a correlation of .57 with the average grade at the end of the first year and .52 at the end of the second; the English test resulted in a correlation with average grades in the two years of .45 and .37, respectively. It seems safe to say, was the conclusion, that a good mathematics test has selective value for engineering students.

Bingham (7) in his Engineering study reported that the score on the mathematical section of the Aptitude Test is one of the most dependable measures for predicting engineering success.

Wilson and Hodges (117) found that marks in freshman mathematics and freshman mechanical drawing courses correlated .42 and .406, respectively, with marks in engineering courses and the Aptitude Test. With the same criterion, the Otis Advanced Intelligence Scale correlated .382. A multiple correlation of .69 for marks in advanced courses in engineering was obtained with marks in freshman courses, and scores on the Otis test. Intelligence tests, previous scholarship, scores on special aptitudes, and achievement tests in science and mathematics seem to provide the best combination for prediction in engineering.

Ruch (88) suggested that the best measures for selection and prediction of students for teacher-training institutions are batteries of tests of general intelligence ability, reading tests, and personality tests. The correlation between the intelligence test and grades for one semester was reported at .80; the correlation between reading test and academic grades was .75; and the correlation between the personality test and grades of special college tests was .72.

Roberts (86) found that the Army Alpha Intelligence Test correlated .75 with the students' average of the first and second semester's grades. Roberts found that success in college is determined by mental ability.

This review of related literature seems to have shown the value of research with this problem of selection and prediction of success of college students as measured by selected criteria. It also seems to indicate that these studies have made many contributions to the problems of vocational success. It would seem highly desirable to conduct continued studies with these matters of selection, prediction, and success.

SECTION II

OPPORTUNITIES AND REQUIREMENTS OF PHYSICAL AND OCCUPATIONAL THERAPISTS

AS A RESULT of World War II, unprecedented development has taken place in both physical and occupational therapy. These two divisions of physical medicine are purposely referred to together, since in modern practice they are inseparably interrelated.

The causes that have been responsible for recent progress in physical medicine are the war with its many physical casualties, a sharp rise in industrial accidents due to the large numbers employed, and the wide-spread publicity given physical medicine along with the stimulus, in the form of educational opportunities and material assistance, furnished by the National Foundation for Infantile Paralysis, the Baruch Committee, and similar organizations. As a consequence of the above causes, in this country more people than ever before have acquired an interest in physical therapy and occupational therapy, have sought to take up these callings, and have come to look upon them as highly desirable and constructive forms of occupational endeavor.

During the cessation of hostilities and the advent of peace, the thought was expressed by some that there probably would occur a more or less sharp decline in the demand for those trained in physical medicine. Although the war has long since ended, there is no evidence to indicate a diminishing need for services in physical therapy and occupational therapy either in the immediate or distant future. For at least another generation the problem of rehabilitation of industrial casualties, war casualties, and other disabilities promise to tax our ability to produce enough adequately trained physical and occupational therapists to meet the continuing employment demand. To be successful with this problem better selection and training of the students must be made.

> "It is becoming more and more apparent that physical medicine is expanding rapidly and that in the future, it will encompass a broad and at present considerably neglected phase of medical practice. As it is now developing, physical medicine is progressing into three major fields: (1) physical therapy, (2) occupational therapy, and (3) reconditioning of the convalescent." (36)

A. Physical Therapists

Definition. —"Physical therapy is the treatment of disease or injury by means of heat, cold, light, scientific massage, therapeutic exercise, water, and electricity." (37) The modalities are used in treatment which is prescribed by a physician for such disorders as sprains, poliomyelitis, fractures, arthritis, and other cases. Physical therapy offers a career for persons interested in physical rehabilitation.

Physical therapy, like occupational therapy, has been discovered to be a tremendous aid in helping the war-wounded for their return to duty or to useful civilian life. Since World War I, it has grown slowly and steadily as its usefulness has been proven in the cases of crippled children, industrial accidents, and infantile paralysis. When the second war came, the need for physical therapists increased considerably.

Outlook. —There were 2000 physical therapists registered with the American Registry of Physical Therapists Technicians in 1946. Less than 100 of them were men.

There was an acute shortage of trained physical therapists during the war and the demand is still greater than the supply. The needs are not being met. Applications for training are now exceeding training capacity so that only the best qualified are selected. This year at the University of Wisconsin, 525 applications were received from forty-four states and Washington, D. C., and five foreign countries from persons interested in enrolling in the Physical Therapy course. Only seventy applicants of the 525 interested were admitted.

Duties. —Working always under the direction of the patient's physician who diagnoses and then prescribes treatment, the physical therapist carries out the doctor's orders. She applies heat and massage to stimulate circulation and to toughen tender or burned skin. The physical therapist uses her fingers and hands in rapid movements. She uses whirlpool baths and ultra-violet lamps to relax and gently exercise muscles which have stiffened or have lost their power of motion. The physiotherapist also employs pulleys, posture chairs, finger extenders, ladders, "walkers," grip strengtheners, exercise machines, and walk ramps to accomplish her purpose.

Qualifications. —What qualifications must a young person have for success in this field? First of all, good technique and sympathetic understanding are essential. She must be a sympathetic person who understands human problems, and yet who is capable of being casual when necessary to help a patient get over any self-pity. A physical therapist should have a keen interest in medicine and in medical problems. Besides this, she must be interested in other people and their troubles, and have an intense desire to be of service to people who are in need.

The physical therapy technician must have a strong interest in her work so that she is constantly motivated. She must possess mechnanical aptitudes, mental ability, judgment, reasoning and verbal ability to such a degree that she will present herself in a self-reliant and successful manner. She must not exhibit any anti-social or withdrawing tendencies. The physical therapist should possess adequate social standards and social skills.

Training. —Special training is absolutely necessary. A registered physiotherapist must have received her education in one of the 26 schools accredited by the American Medical Association. A typical course of study is the one offered by the University of Wisconsin which is approved by the Council of Medical Education, and Hospitals of the American Medical Association, and prepares the student to meet the qualifications and training for registration in the American Registry of Physical Therapy Technicians.

Degree Course in Physical Therapy. —This is a four-year course leading to the degree of Bachelor of Science in the School of Letters and Science and a Certificate in Physical Therapy in the Medical School.

For entrance to this course the student must satisfy the general admission requirements of the University.

Because of the limitation of clinical facilities, at the completion of the sophomore year the highest ranking thirty students will be selected to complete the last two years.

The curriculum requirements are the completion of 124 credits and at least 124 grade points. On the next page is the recommended sequence of the course for the four years and the Certificate Course curriculum.

Certificate Course in Physical Therapy. —The certificate course for physical therapy technicians is twelve months in length and begins in September of each year. College credit toward a degree in Physical Therapy is given for this work.

Candidates for admission must satisfy one of the following requirements:

1. A minimum of 60 college (academic) credits including courses in animal biology and physics.
2. Graduation from an accredited school of nursing.
3. Graduation from an accredited school of physical education.

Courses in chemistry are highly recommended. Due to the limitation of clinical facilities, the number of students admitted to the course is limited to twenty.

Earnings. —Salaries vary considerably, according to the section of the country. A recent survey by the American Physiotherapy Associa-

Freshmen

1st Semester

Chemistry 1a	5
English 1a	3
Language 1a or Social Studies	4
Physics 65	3
Phys. Ed.	0
	15

2nd Semester

Chemistry 1b	5
English 1b	3
Language 1b or Social Studies	4
Physics 65	3
Phys. Ed.	0
	15

Sophomore

1st Semester

English or history	3
Language 10a or Social Studies	3
Speech 1	3
Phys. Ed.	0
P. Chem. 114	3
Electives	3
	15

2nd Semester

English or history	3
Language 10a or Social Studies	3
Psychology 1	4
Phys. Med. 1 (Orientation)	1
Phys. Ed.	0
Zoology 1	5
	16

Junior

1st Semester

*Physiology 115	5
*Psychology 105 (Human Adjustments)	3
*Anatomy 37	6
Home Ec. 119 (Child Development)	2
	16

2nd Semester

Physiology 116 (Applied Physiology)	3
Speech 126 (Advanced Speech Correction)	3
Psychology 106 (Abnormal)	3
Phys. Ed. 56 (Kinesiology)	3
Electives	3-4
	15-16

Senior (12 months)

1st Semester

*Physical Medicine 2 (Preclinical subjects)	1
*Physical Medicine 3 (Medical sciences)	4
*Physical Medicine 4 (Massage)	2
Physical Medicine 100 (Applied anatomy)	3
*Physical Medicine 101 (Pathology)	2
Physical Medicine 103 (Journal Club)	1
Physiology 117 (Special problems in physical medicine)	3
	16

2nd Semester

*Physical Medicine 5 (Therapeutic exercise)	4
*Physical Medicine 6 (electrotherapy)	4
*Physical Medicine 7 (Hydrotherapy)	1
*Physical Medicine 8 (Clinical Practice)	5
*Physical Medicine 102 (Ethics and administration)	1
*Physical Medicine 103 (Journal Club)	1
	16

*Denotes courses taken by Certificate group for the twelve-months course.

tion revealed that the average beginner earns about $3400 a year, but the postwar period is pushing that figure up to $4000. Top salaries are usually around $5400. The therapists with instructional duties have salaries ranging upward to $6000.

Advantages and Disadvantages. —The advantages of a career in physical and occupational therapy are about the same. At present there is a big demand for them both and it will continue. Both have now acquired professional status and are established fields. The work is varied, and offers unlimited chances to develop one's initiative and ability. The hours are regular, usually an eight hour day, five days a week, and a half a day on Saturday. Employment is steady and secure. Best of all, there is the joy of creating comfort and happiness for others, and also the stimulating cooperation of doctors, nurses, social workers, and families.

The postwar future for therapists looks excellent. Physical therapy has never been overcrowded and will probably be used increasingly for various cases.

As for the disadvantages, the therapist must live, in many cases, at institutions which are isolated in remote communities. Educational requirements are constantly being increased, so that the expense of getting the necessary training is sometimes a disadvantage. Therapists often have to work with mental and physical cases which place them in difficult situations, and as a result, are under constant nervous strain. Salaries are not high considering the training required.

B. Occupational Therapists

Definition. —"Occupational therapy is an activity, mental or physical, prescribed by a physician for its remedial value. It is recognized by the medical profession as a valuable adjunct in contributing to and hastening recovery from disease or injury." (74)

"Occupational therapy aims to help the patient in three ways—mentally, physically, and socially. Physically, its function is to increase muscle strength and coordination and joint motion as well as to improve general bodily health. Mentally, its function is to supply as nearly as possible normal activity through avocational projects, and prevocational studies and training. Socially, its function is to arouse interest, courage, confidence, and to give the patient the feeling of social usefulness." (75)

Outlook. —Today there are 2200 Registered Occupational Therapists. There is an increasing need for capable persons in this field. At present there are shortages for clinical, workshop, administrative, and instructional posts. Because of the expansion of veterans' hospitals, as well as civilian hospitals, general increase in the use of this type of treatment for mental patients, crippled children, tubercular patients and convalescents, opportunities will continue to be good.

Duties. —"Occupational therapy is expanding rapidly and seems to be breaking away from its preoccupation with arts and crafts and to be extending into a wider field including new prevocational, avocational, and vocational pursuits. The occupational therapist today is concerned with skills and techniques such as creative arts, recreation and educational subjects." (76)

The occupational therapist must know handicrafts, medicine, and psychology. For each person she must know something of his cultural background, previous interests, and hobbies. She assigns work or projects that will be beneficial, grading it to match the patient's progress and his increasing strength and dexterity. She then measures the results and keeps a chart of the patient's recovery.

The occupational therapist works with mental or orthopedic patients. Some therapists are employed in tuberculosis or children's hospitals or wards. Still others specialize in work with the blind or with patients who are chronically ill. The occupational therapist engages in embroidering, knitting, and basket-making, things in which most women find pleasure. The therapist has such duties as carpentry work, hammering, sawing, painting, whittling, drilling, printing, modeling, crocheting, and making leather goods. She must keep records and provide educational, dramatic, music, and dancing programs.

Qualifications. —"In order to do her job well, the therapist must have certain qualifications. She should have good health, and excellent mental and emotional self-control. The therapist should possess the ability to be casual, with no false cheering up and no overpraise, yet with no minimizing of patient's progress. Extreme patience with cross or fretful individuals, with the mentally ill, and with naughty children is vital." (76)

The occupational therapist must like people and genuinely desire to help them. Tact, poise, a well-balanced personality, initiative to find activities that will awaken a response, and a sense of humor are all essential, as is an interest in medicine and scientific subjects.

The therapist should possess manual dexterity, physical and emotional health, muscle coordination, and special attitudes needed for her work.

Training. —Special training is absolutely necessary. A registered occupational therapist must have received her education in one of the 25 schools accredited by the American Medical Association.

The University of Wisconsin is one of the schools offering a four-year academic course leading to the degree of Bachelor of Science in

Occupational Therapy which is followed by thirty-six weeks of hospital experience in the practice of occupational therapy under the supervision of registered occupational therapists. On successfully completing the course, students receive a certificate from the Medical School as graduate occupational therapists.

During their four years of academic work students are enrolled in the School of Education and concentrate on social and biological sciences together with technical courses in the arts and a variety of manual skills to supply the background required for their clinical work.

For entrance to this course a student must satisfy the general admission requirements of the University. The course is under the joint control of the Schools of Education and Medicine.

The curriculum requirements are the completion of 124 credits and at least 124 grade points.

On the next page is the recommended sequence for the course in Occupational Therapy.

Earnings. —Salaries for beginners range from \$2800 to \$3600 a year. After the occupational therapist has gained experience, she can expect to earn from \$3600 to \$5000 a year. In the executive positions, the salary range is \$6000 to \$8000 per year. Civil Service therapists receive from \$3600 to \$4800 a year.

In the profession as a whole, the salary situation is moving slowly. The rewards, however, are not usually in money, but are in the emotional satisfaction of helping people overcome their disabilities.

Advantages and Disadvantages. —The advantages and disadvantages for the occupational therapist are practically the same as for the physical therapist. There is more opportunity for men in occupational therapy, for the new trend in rehabilitation, a trend which requires that the occupational therapist be able to teach metal trades, woodwork, carpentry, etc., to men patients, may result in an increase in the demand for male therapists.

The lack of training facilities for male students plus the fact that salaries, while they are considered fairly good for women, are rather inadequate for a man to raise a family on, makes the field a difficult one for them!

The opportunities and requirements in physical and occupational therapy are practically similar. There is a great demand today for both types of therapists, and both divisions of physical medicine are expanding rapidly. Since World War I much more training is required of trainees in physical and occupational therapy. Certain pre-professional and personal qualifications are necessary for the students in therapy training. Salaries for the graduate therapists are somewhat attractive and their future is bright as there is a need for therapists which will exist for some time. The advantages for the physical and occupational therapist overweigh the disadvantages.

SECTION III

METHODS AND MATERIALS OF INVESTIGATION

A. Methods of Investigation

FOLLOWING a careful survey of available information from published reports and references, the data presented in this study were secured on the University of Wisconsin campus by using a battery of tests, a rating scale, a data form, and other information gathered from the subjects and the Registrar's office. The University Student Counseling Center also cooperated by furnishing some supplementary data which were not available in the Registrar's office.

The survey of literature included books and pamphlets listed in the Education Index, the Reader's Guide to Periodical Literature, and bibliographies secured from various books and selected references from the American Medical Association, National Education Association, and Encyclopedia of Educational Research.

An especially helpful selective bibliography on Physical Medicine was obtained from Miss Caroline G. Thompson and Miss Margaret A. Kohli, Technical Directors of the Departments of Occupational Therapy and Physical Therapy, respectively.

In this investigation a battery of twelve carefully selected tests covering six areas of desired information was administered to the 234 physical therapy and occupational therapy students. Two kinds of tests were used in this battery, namely, the paper and pencil type, and the performance type. The former was administered to groups, and the latter was given to the students individually.

The rating scale used in this research study was constructed by the Department of Physical Medicine. This instrument is a five-point scale which has three categories: professional preparedness, professional qualifications, and personal qualifications. Five instructors in Physical Therapy, as well as five instructors in Occupational Therapy rated their respective students. The ratings of the five instructors were then pooled and an average for each quality or trait was submitted from the five ratings. This was done for both Occupational and Physical Therapists.

The two-page data form used in this study was constructed by the author. The information was gathered from each student during a short standardized interview and the results were recorded in the data form during each interview.

Freshmen

1st Semester

English 1a (Freshman English)	3
Foreign Language 1 a	4
Art Ed. 50 (Freehand Draw. Perspec.)	3
Physics 65 or 1	3-5
P. E. 15a (First Aid)	1
Physical Education	0
	14-16

2nd Semester

English 1b (Freshman English)	3
Foreign Language 1b	4
Art Ed. 51 (Freehand Drawing)	3
Zoology 17 or 1	3-5
O. T. 1 (Preclinical Train.)	1
P. E. 48 (Recreational Therapy)	1-2
	15-16

Sophomore

1st Semester

Foreign Language 10a	3
Art Ed. 62 (Creative Design)	3
Art Ed. 54 (Elem. Modeling)	2
Speech 1	2
Sociology 1	3
Electives	3
Phy. Education	0
	16

2nd Semester

Foreign Language 10b	3
Art Ed. 62 (Creative Design)	3
Physiology 3	4
Anatomy 36	6
Phy. Education	0
	16

Junior

1st Semester

Art Ed. 70 (Art Metal)	3
Psychology 1 (Introd.)	4
O. T. 2 (Medical Lectures)	1
Home Ec. 85 (Weaving)	3
Electives	5
	16

2nd Semester

Art Ed. 71 (Pottery)	3
Art Ed. 175 (Crafts for O. T.)	4
P. E. 56 (Kinesiology)	3
Home Ec. 94 (Dec. Textiles)	2
Electives	4
	16

Senior

1st Semester

Psychology 105 (Human Adjustments)	3
O. T. 3 (Surgical Lectures)	1
O. T. 4 (Practical Application of O. T. crafts)	2
Home Ec. 119 (Child Development)	2
Home Ec. 15	2
Electives	6
	16

2nd Semester

Psychology 106 (Abnormal)	3
O. T. 5 (Theory of O. T.)	4
O. T. 6 (Clinical Prac. and Dir. Teach.	1
O. T. 7 (Neurology, psychiatry, communicable diseases)	3
P. E. 176 (Therapeutics)	2
Electives	3
	16

Bureau Ednl. Psy. Research
DAVID [illegible] TRAINING COLLEGE

The data form included items of general information such as name of student, age, sex, year in school and course, race, descent, degrees held, birth data, and the other data.

The data form also included a psychograph of all the tests. This test profile contained the names of the tests administered, the raw scores, and the percentile ranks on all tests taken.

The data from the Registrar's office included the actual grade point average, predicted grade point average, high school rank, and general information concerning the subjects.

B. Testing Program

It was found that the testing program needed for this study required a selection of measures for intelligence, interest, special aptitudes, reasoning and judgment, achievement, and personality. The particular combination of measuring techniques required in this investigation was dictated by the specific purposes to be served by the various tests.

It is necessary that a systematic businesslike procedure be employed by the group in charge when the purpose of the testing program is established. This also holds true for the selection of the tests. In selecting the battery of tests the committee scrutinized the pertinent literature, the actual blanks, and the corresponding manuals relating to all materials under consideration. The committee members, each independent of the others, looked over these materials for the purpose of considering the following data:

1. Title, author, publisher, designated function.
2. Significance.
3. Validity. Does it measure what it purports to measure?
4. Reliability. How accurately and consistently does the test measure?
5. Diagnostic value. Does the test give information as to the trainee's difficulties?
6. Norms. What types are available? Are they representative? What is their range?
7. Administration. Is the test easy to give and score? How much time is required?
8. Pupil Performance. What does the pupil do? What dimensions (speed, difficulty, quality) are measured? How are the scores expressed?
9. Construction of the test. How were the items selected? How was the placing of the items carried out?
10. Manual. Does it contain complete and easily intelligible directions? Does it give norms? Does it suggest uses to which the test results might be put? Does it give data about the test?
11. Costs. The cost was considered in relation to the dependability and usefulness of the results.
12. Mechanical considerations. What are the good and bad points about the typography, makeup, packing, size and complexity of the test, material (wooden, paper, cardboard) of the test, etc. *

"Tests have two major functions: a determination of the needs, abilities, achievement status, interests, intelligence, and personality characteristics;.and an appraisal of the outcomes of learning." (18)

Objectives of Testing Program

The information the committee desired to gain from these tests is found in the answers to the following questions:

1. What mental ability does each student possess?
2. Does the student possess the ability to memorize verbal and visual materials?
3. What are the student's interests?
4. To what extent does each student's memory, mental maturity, judgment, reasoning, verbal, and performance abilities limit her?
5. What ability does the trainee have to use her hands and fingers skillfully and dextrously?
6. What ability does the student have to train her muscular movements?
7. Does the student use her abilities and skills in a self-reliant and successful manner?
8. What ability does she have to visualize patterns and relationships without the necessity of preparing drawings of all these relationships? Associated with this is the objective of measuring the student's appreciation of artistic design.
9. What is the student's knowledge of word meanings—both general and scientific vocabulary?
10. What is the student's ability to read with understanding the type of material that she will be expected to read in Physical Medicine? At what rate will she read this?
11. Does she possess social maturation? This is associated with practical judgment.
12. How is her personality?
    a. Does she exhibit any anti-social or withdrawing tendencies?
    b. What deviations, if any, does she possess from the normal personality?
    c. Does she possess adequate social standards and social skills?
    d. Is she maladjusted?

The emphasis on aptitudes, interests, and personality measures is necessary since the

---

*Unpublished paper by John W. M. Rothney, Department of Education, University of Wisconsin, 1947.

majority of students have had little or no experience in clinical work of any sort prior to entering the Physical Medicine program.

The committee after a careful selection of the tests categorized them into six classifications or areas. The areas and tests decided upon are as follows:

Intelligence:	The Wechsler-Bellevue Intelligence Scale and the California Mental Maturity Test
Achievement:	The Wide Range Achievement Test and the Michigan Speed of Reading Test
Interest:	The Kuder Preference Record
Aptitudes:	The Purdue Pegboard, The Minnesota Rate of Manipulation, The Minnesota Paper Form Board, and the Bennett Mechanical Comprehension Test
Personality:	The Minnesota Multiphasic Personality Inventory, and the Washburne Social Adjustment Inventory
Practical Judgment:	The Cardall Practical Judgment Test

The committee recognized that some of these abilities are extremely difficult to measure and for that reason considerable time may be required before adequate measuring devices can be designed and developed for them. The fundamental objective of this aptitude testing program, however, is the extent to which the therapy student is likely to succeed in her professional preparation and later employment. Therefore, the tests to be used were expected to indicate some promise of future success in the practice of occupational or physical therapy. Of course, as stated previously, nowhere is the claim made that these or any other aptitude tests can be used as the only criterion for the admission of students into the Department of Physical Medicine.

Before and after the committee selected its tests much consideration and discussion were given to the specific purposes to be served by the areas of the test battery. Significant objectives of these areas and reasons for their use were discussed in connection with the services which they render in diagnosing, appraising, and guiding. The areas of the battery of tests which were considered relative to their effectiveness in the attainment of the objectives established were intelligence tests, achievement tests, interest tests, special aptitudes, personality and social adjustment tests, and judgment and reasoning tests.

A discussion of the prime considerations and salient functions of each of these areas follows.

Intelligence Tests

"The exact nature of the combination of abilities known as intelligence is not well understood. However, it is definitely known that individuals differ widely in the amount, and perhaps the quality, of it they possess, and that within limits it can be measured." (38)

Many definitions of intelligence have been given. Two of the definitions most commonly quoted are:

Binet:

> "Intelligence is the tendency of thought to take and maintain a definite direction, the capacity to make adaptations for the purpose of attaining the desired end, and the power of self-criticism." (38)

Henmon:

> "Intelligence involves two factors—the capacity for knowledge and knowledge possessed." (38)

There are two types of general intelligence tests, the group and the individual types. Individual intelligence examinations constitute the most accurate devices for the measurement of intelligence. Individual intelligence tests are patterned upon the Binet-Simon tests brought out in France from 1905 to 1911. The Wechsler-Bellevue is rated one of the best of its type today.

Group intelligence tests originated in America during the First World War. The Army Alpha and the Army Beta tests, the latter really a performance scale, were developed for use in selecting personnel for officers' training and for other positions requiring high intelligence.

One of the best group intelligence tests is the California Test of Mental Maturity. Mental ages and intelligence quotients can be obtained separately for the language and non-language sections and for the instrument as a whole.

The intelligence and mental maturity tests bring out the individual's language and non-language abilities. They also indicate the stage of development of such mental processes as memory, reasoning, vocabulary, and spatial relationships. These tests should be interpreted in the light of the individual's background, environment, and training.

Clinical examiners who are relatively satisfied with some form of the Binet-Simon scale will welcome a test, like the Wechsler-Bellevue, which is intrinsically better adapted to adult interests and which is adequately standardized for adults.

Achievement Tests

This type of testing is probably the most straightforward and easily understandable type. The range of achievement measures is and should be greater than a classroom examination.

> "We are familiar with the average grade and the specific pattern of an individual's grades, from which we can judge him to be best in one subject and poorest in another, or better in one type of subject than he is in another. These are important measures for intra-individual and inter-individual analysis, so long as we have all our individuals within one school. But when students move outside the walls, our grades are not always such good measures and not always the only measures predictive of the later adjustment of individuals." (11)

Few psychological studies are complete without school achivement tests. Recent clinical experiences indicate that a thorough examination of a person's personality should include not only interest and intelligence tests, but also tests of the school fundamentals. The correlation between intelligence and school achievement is moderately positive, but not as high as was believed twenty years ago. This finding is consistent with the view that failure in school and in life adjustments may be caused by many factors besides lack of intelligence. Thus it happens that a child of average or even superior intelligence may have great difficulties in learning to read or to figure, while a student of inferior intelligence may become relatively proficient in both reading and arithmetic.

The Wide Range Achievement is a fine diagnostic measure and it is very revealing in the study of individuals.

A need for more reliable tests for speed of reading has been felt both in researches concerning the legibility of type and studies of the improvement of reading ability. Reading disability is practically always a deficiency in the mastery of mechanics and not of comprehension. The Michigan Speed of Reading Test clearly points this out. This is one of the primary reasons this test was selected.

Interest Tests

Interest has long been considered a factor of successful adjustment in any vocation. The possibility that interests adequately measured could be of some prognostic value has resulted in much experimental work on this problem. Various methods have been used for discovering interests. The easiest and most common procedure, and most unsatisfactory, is to ask individuals what they are interested in. This method is extremely inaccurate both from the point of view of recall and the inability to express desire in words. Attempts to develop objective measures of interests have resulted in several effective techniques. Of these the most valuable are the questionnaires or interest inventories.

Kuder (55) and others studied the constellation of interest because it had been demonstrated that one interest isolated from other factors is not reliable for prediction or success or valuable for guidance. The Kuder Preference Record is intended for use in obtaining measures of motivation in various lines of study and of work through a systematic record of a variety of activities.

The preliminary studies disclosed a marked relation between patterns of scores and occupational choïce. More recent research by Traxler and McCall (108) and Yum (120) bolsters these conclusions. Traxler and McCall found that the retest reliability of the Kuder Preference Record was rather high; that the results did not seem to be influenced by practice; that it had high predictive value; and that there were noteworthy sex differences.

Special Aptitudes

> "Aptitudes may be defined as potentialities which can be trained into special skills useful in later job adjustments. For example, clerical aptitude may be the potentiality which could be trained into bookkeeping or accounting or filing skills. Mechanical aptitude might under training become proficiency or skill in punch press operation, welding, and other mechanical trades." (7)

> "An aptitude is the condition or set of characteristics regarded as symptomatic of an individual's ability to acquire with training some specified knowledge, skill, or set of responses such as the ability to produce music, etc." (85)

Aptitudes may be either acquired or inherited. "Present traits" of the individual include not only his ability also his interests, adjustments, character, physical aspects, and all his other aspects, but in evaluating his special aptitudes we usually concentrate on the one, two, three or more aspects which are most valuable for predicting success.

Aptitudes are similar to other aspects of individuals in that they are not all equally strong within the same person, and in that people differ from one another with respect to their possession of any single aptitude.

Like general mental ability, aptitudes may be considered to be relatively stable. It is often far more economical to fit an individual's field of endeavor, either educational or vocational, to his present set of aptitudes than to attempt to change his aptitudes through some kind of training so as to enable success in some field of endeavor.

This does not mean that in evaluating aptitudes we must search for aspects of the student which have not been influenced by experience and training. That is, even aspects that have been subjected to much training—in other words, the present achievements of the students—should be evaluated insofar as these achievements may be considered predictive of future achievement.

In dealing with special aptitudes it is necessary to determine the degree to which various occupations require each of the aptitudes, then to determine the degree to which an individual possesses each of the primary aptitudes, and finally to determine the individual's fitness for the occupation.

The more traditional approach makes the breakdown first in terms of fields of endeavor such as music, art, engineering, law, occupational therapy, physical therapy, and then constructs measuring instruments specifically for the prediction of success in each of these fields. That is, the construction of most tests designed to help in estimating the probabilities that a person will be able to follow successfully an occupation or profession she is considering has attempted to break down mental ability in terms of culturally determined fields, e.g., musical aptitudes, art aptitudes, verbal aptitudes, mathematical aptitudes, nursing aptitudes, legal aptitudes, teaching aptitudes, engineering aptitudes, mechanical aptitudes, clerical aptitudes, and the like.

It is sometimes necessary and advisable to measure aptitudes directly instead of placing the student into a competitive where the end achievement is used as an index of original aptitude. For this reason the committee recommended that the total population be tested with the aptitude tests which will follow in this discussion. There are more mechanical aptitude tests than any other type.

Personality and Social Adjustment Tests

Attempts to judge and predict general personality, character, or the ability of the individual to make personal and social adjustments, are probably more numerous than for any other human characteristics or area of living. Unfortunately science's contributions have not been very fruitful in this area. Within recent years, there has been much interest in the measures of personality and social adjustment. The results give some promise from the standpoint of prognosis and gradually these tests will be thoroughly enough tested to be used in individual prediction. Great care needs to be taken in applying any such measures. The interpretations can be made only by individuals adequately trained.

Paper-and-pencil and performance tests have appeared in great numbers.

Personality and adjustment inventories range from attempts to measure such personality traits as feelings of inferiority, schizoid characteristics, attitudes toward family, life, and school, neurasthenia, etc.

Generally speaking, there are four standard methods of measuring or evaluating personality characteristics and personality adjustment:

1. Rating scale on which outside observers rate the student's possession of personality traits.
2. Intensive interviewing in which skilled workers (psychologists, psychiatrists, social workers, clinical workers) use the interview as a tool to probe the personal adjustments and personality characteristics of an individual.
3. Anecdotal records or behavior journal which represent fruitful information for measuring or estimating personality.
4. Paper-and-pencil tests of personality which student fills out himself.

The scores on a personality test can usually be considered to tell only that a maladjustment does exist, or the general area of the maladjustment, rather than *what* the maladjustment is. Since there are no right or wrong answers in a personality test the interpretation of personality test scores becomes a matter of importance. The highest score need not necessarily be the best score or the best adjusted student may be the one who makes about an average score on a personality test. The counselors or clinician should study carefully the manual of interpretations which accompanies the personality test.

Today one of the reasons for sometimes being hesitant about adopting or accepting psychological tests is that we often think of tests in terms of intelligence, aptitude, and dexterity, and these tests do not, of course, take into account the more general personality traits of the individual. An individual might be very high in mental ability and possess strong aptitudes and interests and yet have a personality that would not only make him unfitted for the occupation for which he is interested but would also make him a definitely undesirable individual to employ in any capacity. Bad personality characteristics and maladjustment can make it difficult for anyone to work on any job. It is for this reason that personality tests and scales are being developed and used to measure incipient personality maladjustments and temperament traits. In this study two reliable and valid personality measures were used, the Minnesota Multiphasic Personality Inventory and the Washburne Social Adjustment Inventory.

Judgment and Reasoning Tests

Under reasoning and judgment tests are included tests of inference and of problem solving. In inference tests the subject is asked either to detect inconsistencies or to make inferences of his own. Greene (35) reports that one of the better judgment or reasoning tests is the one which the subject is asked to check the inferences which may be correctly drawn from an account of a situation where there are four or five variables involved. The unit of measurement in these tests is usually a judgment of true, false, or undetermined relationship.

What is practical judgment? This question was asked of a number of psychologists and guidance personnel. The inquiry resulted in the consensus that practical judgment and its corollaries, on-the-job judgment and common sense, consists of the recognition of possible alternatives of action and the ability to select the best.

There are very few judgment tests. The most accepted scale for practical judgment is Cardall's, the instrument used in this research study. It has items covering social and business situations and those common to a normal environment. The Cardall Practical Judgment Test gives a measure of social maturation which is of material importance in therapy work.

The twelve tests (covering the six areas of desired information) included in this battery are described and discussed in the following pages.

The Wechsler-Bellevue Intelligence Scale

Author: David Wechsler

Purpose:

The primary purpose of this test is to give a valid and reliable measure of the subject's global intellectual capacity and to gather data regarding the testee's mode of reaction, his special abilities or disabilities, and some indications of his personality traits.

Description:

This one-hour test is administered individually. The most obviously useful feature of the Wechsler-Bellevue Scales is their division into a verbal and performance part. Its a priori value is that it makes possible a comparison between a subject's facility in using words and symbols and his ability to manipulate objects, and to perceive visual patterns. It yields a measure of general intelligence that may be expressed in terms of mental age and intelligence quotient.

The verbal parts of the test are general information, general comprehension, arithmetical reasoning, digits forward and backward, similarities, and the vocabulary test. The five performance, or non-verbal parts, are picture completion, picture arrangement, object assembly, block design, and the digit symbol test. Therefore, the Bellevue Intelligence Examination consists of eleven tests, six of which are described as verbal and five as non-verbal, or performance. Wherever it is practicable it is desirable to give the Full Scale. The vocabulary test is an alternate scale.

The information test requires knowledge which one picks up in everyday life as well as in school and in pursuing cultural interests. The total information score and inspection of single items will tell something of the subject's endowment, educational and cultural opportunities, and interests.

The comprehension test is a measure of the subject's soundness of judgment, that is, of his ability to make use of his intellectual assets in a manner not only logically correct but emotionally relevant.

The arithmetic test involves voluntary concentration. The test involves a focusing of consciousness upon a problem.

The digit span test is considered to be a measure of efficiency of attention, of the degree of freedom of receptivity.

The similarities test gives a measure of verbal concept formation. It is a test where the subject has to describe in what way various pairs of named objects are alike.

The vocabulary test is helpful in estimating the original endowment of the subject and thereby detecting impairments in other functions.

The picture completion test (in which a missing part in each series of pictures is to be pointed out) requires utilization of knowledge and involves voluntary concentration.

The picture arrangement test measures the subject's planning ability which is intimately related to the function of judgment. In this test a set of pictures have to be placed in a certain order so that they tell a sensible story.

The object assembly test, a jig-saw type of test, taps the ability to analyze into its parts a given pattern in order to be able to reconstruct it from its parts.

The block design test, where drawn designs have to be duplicated with blocks, measures the ability to be guided by visual impression in motor action of the hand in reproducing a pattern. This is required for good performance.

The digit symbol test in which, according to a given sample, numbered empty squares have to be filled by symbols corresponding to the numbers is closely related to the block design and object assembly tests. It measures the ability to reproduce a pattern.

Norms:

Different age norms for all ages from 7 to 70 years have been established on a total of 3499 cases.

Scoring:

Successes or the raw score with each type of content are translated into weighted scores. By means of tables these weighted scores can singly be equated to Mental Age and Intelligence Quotient and then into percentile values. This is done for the Verbal and Performance Scales, as well as for the Full Scale (Verbal score plus the Performance score).

Validity and Reliability:

A correlational study between the Wechsler-Bellevue Scales' I. Q.'s and the I. Q.'s obtained by the Stanford-Binet Test was made. The correlation for 75 cases (ages 14-16) was .82. This is about as high as most other accepted tests of intelligence correlate with each other. The results show that the Wechsler-Bellevue Scales measure about the same thing as Binet.

Balinsky, Israel, and Wechsler (3) found the following correlations: Bellevue I. Q.'s with Psychiatrist's recommendation, r = .70, and Bellevue I. Q.'s with the Stanford-Binet, .79. Rabin (84) reported a correlation of .74 of the Wechsler-Bellevue Full Scale and the Army Alpha. Wolfson (118) found an .81 correlation of the Bellevue Test with the Henmon-Nelson Examination.

The reliability coefficient of Information, Digit Span, Picture Completion, and Block Design against Comprehension, Arithmetic, Picture Arrangement, and Digit Symbol was .90. The reliability of the Verbal I. Q.'s against the Performance I. Q.'s was .83.

California Test of Mental Maturity—Advanced Series

Authors: E. T. Sullivan, W. W. Clark, E. W. Tiegs

Purpose:

Its primary purpose is to make for each student a diagnostic evaluation of those mental abilities which are related to, or determine, her success in various types of school activity, in order that the teacher, counselor, and school administrator may utilize this information directly in aiding students. The test provides a diagnostic profile designed to show graphically the relative extent to which each student possesses these abilities, thus the counselor may provide to the maximum the guidance which such a profile may suggest.

The test is designed to be diagnostic and analytical and yields three mental ages—a non-language mental age and intelligence quotient, a language M. A. and I. Q., and the mental age and intelligence quotient characteristics of the familiar intelligence test.

Description of Tests:

These tests use language and non-language situations in testing memory and attention, orientation in spatial relations, reasoning abilities and vocabulary. Tests of visual acuity, auditory acuity, and motor coordination are given, but serve merely as pre-tests.

The four tests and their component parts are:

A. Memory
   1. Immediate Recall
   2. Delayed Recall

B. Spatial Relationships
   1. Sensing right and left
   2. Manipulation of areas
   3. Foresight in spatial situations

C. Reasoning
   1. Opposites
   2. Similarities
   3. Analogies
   4. Number Series
   5. Numerical Quantity
   6. Numerical Quality
   7. Inference

D. Vocabulary

"Each test is built upon a plan which (1) takes into recognized mental processes or functions, and (2) corresponds with present knowledge of the nature of mental growth and development."

Reliability:

"Needless to say, the number and variety of test situations assures a high reliability. The test data for 600 pupils in grades 7 to 12 have been analyzed and reliability coefficients have been computed by the split-halves method and corrected by the Spearman-Brown formula. The reliability of the Total Mental Factors score is .96; Language Factors, .95; Non-Language Factors, .93; Memory, .92; Spatial Relations, .88; Reasoning, .90; and Vocabulary is .94." (12)

Validity:

The validity of this test, like any mental test, is difficult to establish. "There are no purely objective criteria or standards which correspond to the factors or abilities in terms of which conceptions of mentality are currently discussed. The traditional method of correlating the results of this battery with the averages of several other intelligence tests reveals that the Total Mental Factors I. Q.'s obtained with this test may be used for comparative purposes with other intelligence tests. Further evidence of validity must await further knowledge of the nature of mental maturity." (12)

Norms:

The norms are comparable to those regularly obtained by use of individual psychological examinations and well-standardized group tests. The norms were established on 600 students in grades 7 to 12.

Wide Range Achievement Tests

Authors: Joseph Jastak and Sidney Bijou

This test measures separately oral reading, spelling, and arithmetical computation. It is designed for clinical, remedial and vocational studies of individual children and adults. The application of the Wide Range Achievement Tests requires no specialized training. If the Manual of Directions is carefully studied and the directions are faithfully observed, the tests may be accurately administered. A discussion of the three parts of the four-page test follows.

Reading:

This part of the test contains 120 words which are pronounced by the student to the examiner. The subject reads the words across the page. The first clear-cut response should always be accepted and scored as either right or wrong. The raw score is the number of words correctly pronounced.

Reliability and Validity:

The test is one of the most reliable and valid tests of reading as a motor skill. Besides, its administration and scoring hardly takes more than five minutes. Its correlation with the new Stanford Paragraph Reading Test is .81, and with the New Stanford Word Reading Test is .84. This is based on 389 cases. The reliability coefficient determined by repeated testing was .95 (110 cases).

Norms:

Grade ratings, mental age equivalents and test quotients are based on Terman-Merrill's table of I. Q.'s which are listed in the table of the manual. The norms are based on 1492 students from first grade up to the senior year in college.

Spelling:

The first page of the test blank is reserved for the spelling test. Three columns of 23 lines each furnish room for 69 words. Since the reading test is given first, its grade rating serves as a convenient clue to the spelling range to be administered. The words are dictated individually to the subject and he, in turn, writes each word on the blank reserved for the words. The test may begin with a word about two or three grade levels below the subject's reading grade. The dictation is discontinued on reaching the tenth consecutive misspelling in individual testing. In group testing the number of words dictated should be larger to allow for the greater variability of the individuals in the group. The raw score is the number of words correctly spelled plus seven. If a student gets ten words right, her score is ten plus seven.

Norms:

They are established on 4052 cases. Grade ratings, mental age equivalents and test quotients are based on Terman-Merrill's table of I. Q.'s which are listed in the table of the Manual.

Reliability and Validity:

The average time of administration is eight minutes. Its correlation with the New Stanford Dictation Test is .93. The reliability coefficient determined by repeated testing was .88 (120 cases).

Arithmetic:

This test consists of 79 computation problems printed on the inside pages of the test blank. Additions, subtractions, multiplications, and divisions are printed in separate rows to eliminate errors due to change in the operations. The test may be administered individually or in a group. There are no specific standard instructions. The subject is simply asked to answer or solve as many of the problems as he can.

The total arithmetic score is the number of problems correctly solved. The grade ratings corresponding to the raw scores are found with the mental age equivalents and test quotients in the table of the manual.

Reliability and Validity:

The test was given to 4052 students from first grade up through the senior year of college. It correlates with the New Stanford Arithmetic Computation Test to the extent of .91. The reliability coefficient determined by repeated testing was .90 (120 cases).

Michigan Speed of Reading Test, Form II

Author: Edward B. Greene

Purpose and Description:

This test is designed to measure one's ability and speed of reading. This test has two forms, I and II; the latter is used in this study. Each form consists of seventy-five 30-word sections of similar complexity. In each section there is one word in the last part which spoils the meaning of the first part. The task is to find the "wrong word" and cross it out. In order to do this it is necessary to read nearly the whole paragraph, for each part alone makes sense and it is only by comparing the two halves that the "wrong word" is found.

"The words in both forms, I and II, were chosen with few exceptions from the first 5000 of Thorndike's Teacher's Word Book. The subject-matter is taken from school and home situations common to third and fourth grade children. Finally, from 300 items, 150 were separated into two piles in a random order and printed in identical fashion." (39)

Administration:

The forms can be administered under group conditions to students from third grade to college seniors. Seven minutes are allowed for each form. Before administering the test for record allow the students to work the sample sections.

Norms:

The norms show the subject's position in his grade in terms of a standard score and centiles. The equivalent of the raw score is also given for average ages of persons in grades, for I. Q. distributions, and for Usual Letter grades. The norms are based on 3302 persons ranging in age 8 to 20 years.

Reliability and Validity:

The reliability of Form II by split-halves method is .86, and for Form I the reliability coefficient is .83.

The Kuder Preference Record

Author: G. Frederic Kuder

Description:

The Kuder Preference Record consists of 330 paired-comparison items arranged and scored in such a manner that a profile of preference scores is obtained in the following nine general areas: (1) scientific, (2) computational, (3) mechanical, (4) persuasive, (5) artistic, (6) literary, (7) musical, (8) social service, and (9) clerical. The Preference Record is designed for use with both high school and college students. There is no time limit, but most persons can finish it within a class period of forty minutes.

The unique format of the Preference Record is worthy of comment. The test itself consists of a direction page and eleven pages of test items which are attached permanently by means of ten small rings within a heavy cardboard cover. A detachable answer sheet containing eleven columns corresponding to the eleven pages of the test rests inside the back cover. The test booklet may be used repeatedly, therefore, used answer sheets may be removed and new ones inserted. There are three types of answer sheets—one for hand scoring, one for machine scoring, and one for self-scoring. The self-scoring feature appeals to schools with limited clerical resources.

Validity:

"Kuder is to be commended in that he has provided more statistical data concerning his Preference Record than usually is available for a new test. He has shown that the preference type of item is fairly stable in that it measures about the same thing when set up in different form and text. He has also shown that when academic aptitude scores and average college grades were used as criteria of validity, the validity of items originally selected in rank order form will, when tried out with a different group, hold up about as well in paired-comparison form as in the original form." (11)

From research, the results of the Kuder Preference Record indicate in general that the names assigned to the various scales are appropriate in terms of the type of occupation entered as well as in terms of the activities for which the scale is scored. Chemists are found to be particularly high on the scientific scale, writers on the literary scale, musicians on the musical scale, accountants on the computational scale, and so on.

Lehman (59) studied graduates of the School of Home Economics of Ohio State University who have entered various vocations. The home economics teachers scored relatively high in social service, artistic, and scientific fields, and low in the persuasive and clerical fields. Hospital dietitians were very high in the scientific and social service areas and somewhat high in the computational field, while being extremely low in the persuasive area.

Hahn and Williams (41) studied the Kuder Preference Record scores of women reservists in the Marine Corps engaged in various military occupations. Among the significant deviations from the average were the following: high mechanical, artistic, and scientific preferences, and low literary, musical and clerical preferences for aviation assembly and repair workers; high clerical and low mechanical preferences for stenographers; high scientific and mechanical preferences for aviation machinist mates; high persuasive and social service preferences and low scientific preferences for duty non-commissioned officers; and high mechanical and artistic preferences and low musical preferences for officers, technical specialty.

Mangold (66) and the author of the test made a comprehensive study of the profiles of women students preparing for various occupations such as physician, nurse, occupational therapist, dental hygienist, and so on. They found among seventy occupational therapists high mechanical, scientific, artistic, and social service scores, and low computational and clerical preferences.

Reliability:

The reliabilities obtained for the Kuder Preference Record scales from various groups are all close to .90. Traxler (108), with forty-one graduate students had the following reliabilities (using repeat reliability with a three day interval): mechanical .97, computational .98, scientific .95, persuasive .96, artistic .97, literary .95, musical .95, social service .93, and clerical .98. Triggs (109), reports that 166 college students had the following reliabilities: mechanical .94, computational .90, scientific .93, persuasive .93, artistic .91, literary .90, musical .90, social service .91, and clerical .89.

Mangold (66), in a similar study of groups in various college majors, obtained results similar to those of Triggs and Traxler. The median of his four groups was .91 and the average reliabilities for the different scales were close to .90.

Norms:

The profile sheet for use in interpreting scores has been developed from distributions of scores obtained from 1858 high school boys in the sophomore, junior and senior classes, and from 2005 high school girls in the sophomore, junior, and senior classes. In each case the three classes were weighted equally in computing norms. The profile sheet contains separate profiles for boys and girls.

Kuder (55) reports norms for college students are being developed, although there is some question as to the value of general college norms in view of the fairly wide variation in average scores among groups from different colleges.

The Purdue Pegboard

Author: Developed by Purdue Research Foundation

Purpose:

"The Purdue Pegboard is a test of manipulative dexterity designed to assist in the selection of employees in industrial jobs requiring manipulative dexterity, such as assembly, packing, operation of certain machines, and other routine manual jobs of an exacting nature." (83)

The test provides separate measurements of the right hand, left hand, and both hands together, and measures dexterity for two types of activity. One type of activity involves gross movements of the hand, fingers, and arms, and the other involves primarily what might be called "tip of the finger" dexterity needed in small assembly work.

Description:

The pegboard is a wooden board with two rows of holes (25 in each row), equipped with four compartments (cups) at the top end which hold pins, washers, and collars. The extreme right and left hand cups each contain 25 pins. The pocket (cup) immediately to the right of the center has 20 collars, and the compartment immediately to the left of the center contains 40 washers.

Administration:

The subject should be seated comfortably at a table approximately 30 inches high with the pegboard directly away from the subject with the cups containing the assembly parts at the far end of the board.

The first sequence tests the right hand. The subject is instructed to pick up one pin at a time from the right hand cup and place these pins in the right hand row, starting with the top hole. The testee is allowed to put in five or six pins for practice before this part of the test is begun. The subject is allowed 30 seconds to put in as many pins as possible with the right hand. The first score recorded is the total number of pins the subject places with the right hand.

The second sequence tests the left hand. The same procedure as above is followed. The subject is allowed 30 seconds to take pins one at a time from the left hand cup and place them in the left hand row of holes starting with the top hole. After the right and left hand sequences have been administered, all pins are returned by the subject to the right and left cups, respectively.

The third sequence tests both hands simultaneously. The subject takes pins from the right cup with the right hand and from the left cup with the left hand and simultaneously places the pins in the two rows of holes. Another 30 seconds is allowed for this sequence. The score recorded is the total number of pairs of pins placed in the half minute of testing. After this sequence is completed, a fourth score is obtained by totaling the scores the subject has made on the first three sequences.

The fourth sequence tests more minute finger dexterity and consists of assémbling the pins. collars, and washers. In this sequence the subject is given one minute to make as many assemblies as she can with the pins, collars, and washers. The most important point to explain is that both hands should be operating all the time alternately, one picking up a pin, one a washer, one a collar, and so on.

Scoring:

The score for the right hand is the number of pins placed during the thirty-second testing period. The score for the left hand pin test is the number of pins placed during the thirty-second period, and for the two hand pin test, the number of pairs of pins placed during the half minute test period. The score on the assembly test is the number of assemblies made during the one minute testing period.

The test thus results in five scores: a right hand score, a left hand score, a score for both right and left hands used simultaneously, a combined score of the first three sequences, and a pin-washer-collar-washer "assembly" score.

Norms:

Separate norms are available for each of these five scores which give raw score norms and their percentile equivalents for one trial, two-trial, and three-trial administrations. The one-trial norms are based on 954 applicants.

Reliability and Validity:

The reliability of the test on a one-trial basis for the right hand, left hand, and both hands is .62, and for the assembly test is .72. The two-trial method yields reliabilities of .77 and .84, respectively. For the three-trial method the reliabilities are .83 and .89, respectively.

As to the validity of this test, considerable data have been accumulated, but government regulations prevent the release of these data. The test was standardized in numerous government defense plants on a wide variety of industrial jobs. In one plant more than 4000 employees were tested, those making a score in the upper 30 percent being placed on the jobs requiring greatest dexterity. This procedure has given thoroughly satisfactory results in employee job performance.

The Minnesota Rate of Manipulation Test

Author: W. A. Ziegler

The Minnesota Rate of Manipulation Test is designed to measure native speed capacity or muscular reaction time. In certain occupations, both shop and office, rapid use of hands and fingers in machine operations and handling of tools and materials is such an important factor that a worker's success is largely dependent upon the speed at which she can work. Speed is primarily a native trait, improvable only to a limited extent. When workers increase production per unit of time, the increase is due to improved efficiency technique, and not to an actual increase in physical speed of manipulation. Muscular reaction time is a unique trait the magnitude of which is easily and reliably determined by the Minnesota Rate of Manipulation Test. Any means that will make it possible to avoid useless and costly waste of time in trial and that will dismiss method of selecting applicants for work requiring speed will be a great benefit to the employer and applicant alike. This test measures two kinds of speed, hand speed (placing test) and finger speed (turning test). These abilities are different from finger and tweezer dexterity tests.

Description:

The test is a rectangular board which has 60 round identical holes cut in it (4 rows of 15 each). The subject must place 60 round, equatone, identical blocks into these holes in a prescribed order for the placing test. The subject must pick up the blocks with one hand, turn them over, and place the blocks back into the same holes with the other hand for the turning test. The two ends of the blocks are painted, red on one and yellow on the other, for ease of checking and maximum visibility. Scores for each test are the total times in seconds for four trials.

Norms:

Random samplings of total population as selected by the Minnesota Employment Stabilization Research Institute were used in establishing the norms. The raw scores are transcribed into scores of percentile, quartiles, percent placement, and qualitative terms of speed.

Validity and Reliability:

Correlation between hand and finger speeds is .57. Different average scores are made by known different levels of mechanical ability. A publishing house shifted employees around so that those with high speed scores worked where fast handling of materials was important and the output was doubled. A wholesale house increased its production ratios by effective selections based on test results. Jurgensen (49) tested a group of men hired as converting machine operators in a paper mill and the reliability was .75. The reliability of the test with cashiers at a department store, reports Jurgensen, was .90. The correlation between the Rate of Manipulation Test and the Pennsylvania Bi-Manual Worksample was .46.

The Revised Minnesota Paper Form Board Test

Authors: R. Likert and W. H. Quasha

Description:

The Revised Minnesota Paper Form Board Test is a modification of the paper form board test prepared by a group at the University of Minnesota. The original paper form board test was found by its authors to have a relatively high validity for the prediction and measurement of mechanical ability and to differentiate between the various grades of proficiencies of mechanical ability. The earlier test was difficult to score and the scoring involved an undesirable element of subjectivity. This revision has overcome some of these limitations. It is used because it appears to have important guidance values which may be developed with further use and experimentation. The revision simplifies the scoring by changing the items to the multiple-choice type. The present form has 64 items, each consisting of a diagram of from 2 to 5 disarranged parts of a geometrical figure. In each item, there are 5 diagrams indicating how these parts might appear if fitted together. Only one of the 5 choices is correct. The standard time limit is 20 minutes. Norms are available for time limits of 14 and 25 minutes, but the 20 minute time limit should be used. The test presumably measures the same type of ability as is required in the test of spatial relations. It measures ability to visualize spatial patterns in two dimensions. The evidence accumulated appears to indicate that high scores on this test are predictive of ability to learn mechanical drawing and it is predictive of success in mechanical occupations. The paper-and-pencil task involves more of the linguistic type of intelligence, than does the formboard.

Purpose:

The test can be used for examinees of 9 years and above. The authors have not shown conclusively that the scores for older age groups have the same meaning as those for the younger groups.

Norms:

The authors furnish percentile norms for a considerable number of groups such as freshman engineering students, senior engineering students, liberal arts college freshmen, high school seniors, adults, elementary school boys and girls.

Reliability:

The coefficient of reliability indicates the degree to which the test consistently measures the same factor. The reliability of a single series of the test is 0.85. When both series are administered the reliability is 0.92. It should be noted in this connection that it is more desirable to give both series in preference to one, as the score obtained from giving both series will be more reliable.

Validity:

The coefficient of validity indicates the degree to which the test correlates with a criterion of mechanical ability. Of all the tests used in the Minnesota Mechanical Ability Battery, the Minnesota Paper Form Test was the only paper-and-pencil test to give satisfactory correlations with a criterion of mechanical ability. Engineering students score about 6 points higher on the average than do arts college students. The authors report a correlation of +.49 obtained between scores made on the revised form and mechanical drawing grades for 174 students. There is a correlation of +.40 between the test and Otis Self-Administering Examination.

Mechanical Comprehension Test, Form AA

Author: George K. Bennett

Description:

This test is designed to measure the capacity of an individual to understand various types of physical relationships. This type of ability is important in physics courses, trade school courses and in engineering schools. Also it is designed to measure ability to understand mechanical devices. It consists of sixty drawings and simply phrased questions. It has a minimum effect of environment and rote knowledge of physical laws.

This form is suitable for use with high school students or with adults of comparable education. Probably as a result of cultural factors, girls tend to dislike the type of problem present. Experience and education do have an appreciable influence on scores. However, most of the problems are so presented as to require understanding of the principles involved rather than rote knowledge.

Administration:

The test has no time limit. Ordinarily a great majority complete the test in twenty to twenty-five minutes, and little is gained by allowing more than thirty minutes. The test has a pictorial booklet with the sixty questions which are simply phrased. The answers are placed on an answer sheet which can be machine-scored or hand-scored. In either case, the score equals the number right minus one-half the number wrong.

Interpretation:

The score for each person tested should be evaluated in relation to the norms for an appropriate group. Students about to leave high school may be compared with high school seniors, engineering school freshmen, or candidates for technical or apprentice training courses. For trade and apprentice courses, manual dexterity is important.

Reliability and Validity:

The self-correlation of the test, corrected by the Spearman-Brown formula, is .84 for a single grade. Standard error of a score is 3.7 points.

At this time direct evidence of validity is meagre. Each item correlates with a composite of mechanical tests. Fry (31) reports Arts College freshmen (65 cases) earn a mean score of 40.3 which is significantly lower than the mean (46.2) for first year engineering students. In the case of 315 applicants for technical courses a statement as to previous training in physics was made. For 220 individuals reporting such previous training the mean was 41.7. For 95 persons reporting no training in physics the mean was 39.7.

The correlation of this test with other tests are as follows: Otis Self-Administering Test of Mental Ability .25; Minnesota Paper Form Board .59; Detroit Mechanical Aptitude Examination .54; and Short Alpha Examination .47.

The Minnesota Multiphasic Personality Inventory

Authors: Starke R. Hathaway and J. Charnley McKinley

Purpose:

“*The Minnesota Multiphasic Personality Inventory* is a psychometric instrument designed ultimately to provide, in a single test, scores on all the more important phases of personality. The point of view determining the importance of a trait in this case is that of the clinical or personnel worker who wishes to assay those traits that are commonly characteristic of disabling psychological abnormality.” (45)

Description:

There are two forms of the Inventory—the group (paper and pencil) and the individual (card-sorted) types. The individual type was used in this study. Both forms of the instrument comprise 550 statements, each printed in simple language on a separate card or in the booklet. The statements cover a wide range of subject matter—from the physical condition to the morale and the social attitudes of the individual being tested.

The personality characteristics now in available form for scoring are hypochondriasis, depression, hysteria, psychopathic personality, masculinity-femininity, paranoia, psychasthenia, schizophrenia, and hypomania. Others are in the course of development.

“A brief description of the aforementioned personality characteristics follows:

Hypochondriasis	measures abnormal concern about bodily functions, health, complains of pains and disorders. A high score shows that person is unduly worried about his health.
Depression	measures depth of the clinically recognized symptom or symptom complex, depression. A high score indicates poor morale with a feeling of uselessness, lack of self-confidence.
Hysteria	measures degree to which subject is like patient who has developed conversion-type hysteria symptoms (systemic complaints, paralyses, cramps, gastric or intestinal complaints, or cardiac).

Psychopathic Deviate	measures the similarity of the subject to a group of persons whose main difficulty lies in their absence of deep emotional response. A PD digresses from social mores and breaks out in stealing, lying, alcohol, and sexual immorality.
Interest (Mf)	Measures tendency toward masculinity and/or femininity of interest pattern. Mf score is important in vocational choice.
Paranoia	measures persons characterized by suspiciousness, oversensitivity, delusions of persecution, with or without expansive egotism.
Psychasthenia	measures the similarity of subject to psychiatric patients who are troubled by phobias or compulsive behavior, excessive worry.
Schizophrenia	measures the similarity of subject responses to those patients who are characterized by bizarre, and unusual thoughts or behavior.
Hypomania	measures the personality factor characteristic of persons with marked over-productivity in action and thought. The hypomanic patient has usually gotten into trouble because of undertaking too many things." (44)

Administration:

The subject is asked to sort all the cards into three categories indicated by guide cards, True, False, and Cannot Say. The time required for administration varies widely but is rarely longer than 90 minutes and is commonly as short as 30 minutes. Very little instruction and no supervision are required.

After the subject has taken the test the numbers of cards thus sorted are tabulated on a simplified record sheet, that provides the basic record to be scored by any one of the various keys.

Norms:

The chief normative group for the standardization of the scales consisted of 810 individuals representing a cross section of the Minnesota population as obtained from visitors to the University Hospital. "A break down of the group is 351 married women, 119 single women, 229 married men, 111 single men. The sampling is fairly adequate for the ages 16 to 55 and for both sexes."

Reliability and Validity:

"The chief criterion of excellence has been the valid prediction of clinical cases against the neuropsychiatric staff diagnosis, rather than statistical measures of reliability and validity. Nevertheless, the reliability coefficients of scales so far developed in final form range between .71 and .83." (73)

These correlations were obtained by testing and retesting a group of 40 normals at intervals of a week or less up to several years.

This test has included in the final profile of the Inventory three validating scores. They are the Question score, Lie score, and the Validity (F) score. They are briefly described as follows:

Question	"Consists of the total number of items put in the CANNOT SAY category. The size of the score affects the significance of the other scores. A "borderline score" (Question score) probably means that the subject's actual score, if he had not used the Cannot Say category at all, would deviate farther from the average than his observed score indicates.
Lie	a measure of the degree to which the subject may be attempting to falsify his scores by always choosing the response that places him in the most acceptable light socially. A high Lie score does not invalidate the other scores but indicates that the true values are probably higher than those actually obtained.
Validity	Serves as a check on the validity of the whole record. If the Validity (F score) is high, the other scales are likely to be invalid either because the subject was careless or unable to comprehend the items, or because someone made extensive errors in entering the items on the record sheet." (44)

Washburne Social-Adjustment Inventory

Author: John N. Washburne

Purpose:

The primary purpose of the Washburne Social-Adjustment Inventory is to determine the degree of social and emotional adjustment of an individual. The score is designed to give a separate measure of development in each of six traits, which are very slightly correlated with intelligence and are highly correlated with social and emotional adjustment. The score is also designed to give a measure of adjustment in all the traits combined. In the total social-adjustment score, strength in one trait may compensate for weakness in another, as is also the case in actual social situations. Scores for the separate traits reveal particular difficulties, and this is of service in diagnosis and remediation. The test is a useful instrument in counseling, teaching, grouping, and experimenting.

Description

The test is an eight-page booklet containing the questions and an answer sheet, and student's record. The Inventory has a set of three scoring stencils, or keys, for hand scoring. Separate answer sheets and stencils specially prepared for machine scoring are also available.

The test is self-administering and self-explanatory to the person taking it. It contains one essay-type and six objective-type subtests of social and emotional adjustment. The test also contains a subtest of truthfulness, or test accuracy, the score of which indicates how valid and reliable the total score is. The essay-type subtest (wishes) is of importance chiefly for purposes of interviewing and differentiating between degrees of superior adjustment, and need not be scored in order to compare individual test results with the norms.

The subtests are Truthfulness, Happiness, Alienation, Sympathy, Purpose, Impulse-Judgment, Control, and Wishes. From the initial letters of the subtests—except Wishes—has been coined the word "Thaspic" by which the edition of the test is known.

A brief description of the subtest follows:

Truthfulness: This trait indicates to what degree there was relative freedom from deliberate or unintentional inaccuracies in answering the questions.

Happiness: The term includes contentment, a sense of well-being, and the feeling that life is worthwhile.

Alienation: This term measures the individual's sense of social membership and acceptance, also a sense of basic similarity, or of common humanity with others.

Sympathy: This term is defined as sensitive, empathetic, non-negative responsiveness to people.

Purpose: This personality measure indicates satisfactory adjustment as one definitely directs himself toward his aim, goal, evaluation, selection, and effort.

Impulse-Judgment: This term is defined as the ability to judge well between conflicting impulses, so that satisfactions are not discounted in favor of easier or more immediate but obviously lesser satisfactions.

Control: This trait indicates the sense of self-control, self-regulation, and the ability to make and execute plans.

Validity and Reliability:

The number of populations tested was between thirty and forty and the number of individuals tested was well over 10,000. The bi-serial r coefficient of validity is .90. As a further check on validity, it was found that the correlation between the total scores on the Inventory and intelligence is negligible -.07. Archer (2) found that the coefficient of reliability as determined by a retest of approximately 400 college students after an interval of one semester is .92. The probable error of an individual's total social-adjustment score is 5. For the subtests the approximate of the coefficients of reliability is from .80 to .90.

Interpretation:

There is no time limit for the test, but the average person will finish it in about 45 minutes, including the filling in of the Student's Record. Scores on the Social Adjustment Inventory may be tabulated, and measures of central tendency computed to determine the social adjustment of the group as a whole.

The first step in interpreting the test results is to fill in the Profile Chart. Norms are printed on the face of this chart in order to facilitate graphing the positions of the individuals in relation to the norms. The chart is made merely by picking out from among the printed scores those earned by the examinee in the various subtests and in the total test and then joining these points by a line.

In the middle of the Profile Chart are two heavily leaded lines indicating the median score for junior high school, senior high school, and college students in each of the subtests and in the total test. To interpret the specific scores see the "Percentile Ranks" and the "Levels" printed at the top of each column. The levels are given as Exceptional (Exc.), Superior (Sup.), Well-Adjusted (W-A), Normal (Norm.), Low Normal (L.N.), Borderline (Bor.), and Maladjusted (Mal.).

Test of Practical Judgment

Author: Alfred J. Cardall

This test is designed as a measure of ability to reason in practical situations. It is a test to measure the element of practical judgment in everyday business and social situations. Practical judgment consists of an individual's ability to recognize possible alternatives in solving a problem and to select the best solution. This quality is good in selecting individuals for responsible work. It may be observed that in ordinary situations a number of alternatives spring instantly to mind, the element of judgment operating when one course of action is chosen over others.

"This measurement is unique and statistically independent of such factors as intelligence, academic and social background. Practical judgment is an important consideration in the selection of executives, supervisors, salesmen, and those who are working in contact positions, and those responsible for planning and supervising work. Schools use the test as a measure of social maturation," states Cardall.

Description:

The test is published in a sixteen-page booklet and presents two practice and forty-eight test problems. The majority of the questions require the subject to indicate his first, second, and third choices for solving the problems. The subject punches his responses on the test booklet, and his or her answers are automatically recorded by the pin-punch method as correct or incorrect. The highest possible score is 342 points. The highest score to date is 271, recorded by a traffic officer.

Reliability and Validity:

The reliability of this test, computed on a homogeneous group, is .88 by the method of split-halves and .86 by the Kuder and Richardson t formula. In other samplings, but on a smaller number of cases in which the range of ability has been greater, the reliability coefficient has been materially higher. A number of extensive experiments are being carried on with this instrument in order to determine its predictive significance in a variety of situations. In one study recently completed in a public utility company, test results correlated .56 with "value to the company" in the accounting department as determined by pooled ratings. Executive sampling indicates even higher relationships.

Norms:

The first set of norms is based on 303 high school seniors. School administrators, concerned with social maturation of students or counselees, will find these norms applicable at this level and a year above or below. The other set of norms is based on 212 adults doing supervisory work and is applicable in business and industry.

C. Description of Rating Scale

The simplest method of getting an estimate of personality traits is the rating scale. In the use of this method a scale is first more or less arbitrarily established by the person who is to use it. The extremes represent the most possible or the least possible quantity of the trait, and as many intervening scale intervals can be used as seen desirable.

> "Rating is essentially directed observation. It is of no value unless the observation on which it is based is accurate. Unfortunately, persons confronted with a neat and inviting rating scale are tempted to cooperate by putting checks in the spaces provided for that purpose even though they have only a meager basis for judgment. The mechanics of rating is so easy; accurate judgment so difficult.
>
> "The rating scale is a quantitative method of recording and summarizing observations and judgments based on observation. The field of observation, and therefore the scope of rating, is unlimited. Rating may be directed toward the process of learning as well as to the end-products. Thus the rating scale becomes an important instrument for studying the processes by which personality is modified." (70)

The history of the rating scale movement goes back about forty years. Parker (78) attempted to trace the development of rating scales and make a critical analysis of the scales. He found there has been a steady increase in the number of rating scales and a tendency toward more objective ratings, specialization of raters and objective standards of comparison, all leading to higher reliability and validity of the scales constructed.

May and Hartshorne (67) list three devices for obtaining the judgment of teachers and one for obtaining the opinions of students:

1. The conduct rating scale, which uses phrases or sentences to describe the various lev-

els of behavior.

2. A check list, consisting of descriptive adjectives.

3. Portrait-matching, a verbal picture of a given type of personality.

4. The "guess who" test, expressed in a series of descriptions of a single phrase or sentence.

The rating scale used in this investigation, which is the conduct rating scale type, was for the purpose of getting supervisory ratings of the students in training and therapists in the field (one year after completing the prescribed course), and to determine the relationship, if any, between such ratings and the objective measures of the students and therapists in the field based on the test results, grade point average, and predicted grade average.

This rating scale was constructed by the Department of Physical Medicine at the University of Wisconsin to provide a means of reporting systematically the student's personal and professional qualities and characteristics, as well as professional preparedness traits. The first step in its construction was the requesting of the Physical Medicine faculty at the University of Wisconsin to indicate what personality characteristics they considered important for therapists in the field and in their pre-service training. The faculty members submitted lists of personality characteristics, together with explanatory remarks indicating something of what was meant by the various traits named.

From these returns a master list of four professional-preparedness traits, twelve professional-qualifications characteristics, and fourteen personal-qualifications traits was drawn up. Since there appeared to be an avoidable and undesirable degree of overlapping among the qualities, certain combinations and eliminations were made before the rating scale was finally adopted.

Under the professional-preparedness category the following traits were used:

1. Theory:
   a. Knowledge
   b. Application
2. Techniques:
   a. Knowledge
   b. Application

In the professional-qualifications classification the following characteristics were used:

1. Approach to patient
2. Ability to impart instructions
3. Executive ability (leadership-organization)
4. Initiative (take appropriation on own)
5. Neatness in preparation of work
   a. Economy of materials
   b. Care of supplies and equipment
6. Professional ethics and etiquette
7. Records
   a. Recognize importance
   b. Neatness.
   c. Accuracy
8. Professional interest

The fourteen qualities in the personal-qualifications group used were:

1. Adaptability
2. Emotional stability
3. Poise
4. Sense of humor
5. Tact
6. Judgment
7. Dependability
8. Intellectual curiosity
9. Cooperation with personnel
10. Ability to profit by instruction and criticism
11. Loyalty to department and personnel
12. Punctuality
13. Personal appearance
14. Health

The rating scale used in this study has five units, steps, or intervals. They are superior, above average, average, limited, and unsatisfactory.

Five instructors in Occupational Therapy, as well as five instructors in Physical Therapy rated their respective students. The ratings of the five instructors were then pooled and an average rating for each quality was submitted from the five ratings. This was done for both Occupational and Physical Therapists. For those girls already in the field ratings were used which had been judged by one person, and in some cases, two, three, and four raters judged the rating scale.

Many of the judges or raters made use of the space appropriated for "Remarks" in the lower-fourth of the rating scale. Many clarifying statements were written by the raters concerning the students' traits, or the therapists' qualities in the follow-up. This rating scale is strictly a subjective measure.

Weighting of Scale:

The scale had three general categories, thus equal weight was given to the three classifications. The entire rating scale was given a total weight of three hundred, one hundred for each category. The four traits in the first classification were given values of 25 points each, a total of 100. The next category, professional-qualifications, gave 8. 33 points to each of the twelve traits, a total of 100; and the fourteen qualities in the third classification, personal qualifications, were given values of 7. 14 points each, a total of 100.

The faculty of the Departments of Occupational and Physical Therapy have drawn up an interpretational key for the rating scale used in this study. The interpretational key gives explanatory remarks of what is meant by the various personality characteristics and traits named. The interpretational key follows:

Interpretational Key for the Rating Scale

Theory - Knowledge

S. Thorough understanding of medical background, theory of O. T. and related information. Familiar with diagnoses met in patient contact and their implications. Knows precautions to be observed in practice. Understands treatment aims for each type of patient and how to accomplish them.

Theory - Application

S. Has unusual ability to analyze problems and apply knowledge. Always has a planned program suited to the needs of each patient assigned. Follows this program, and can discuss it and modify it if necessary.

Techniques - Knowledge

S. Has unusual knowledge and understanding of skills and techniques, mastery of all major activities, comprehension of many minor ones. Quick to utilize all opportunities to increase knowledge.

Techniques - Application

S. Has high standards of workmanship and executes with precision and skill. Always modifies techniques to patient disability. Shows originality and resourcefulness.

Approach to Patients

S. Inspires and sustains interest, confidence and cooperation in all types of patients. Does not arouse undesirable response. Treats the patient as an individual and observes and considers his reactions.

Ability to Impart Instructions

S. Always considers the mental, physical and intellectual abilities of the patient. Knows thoroughly what is to be taught and has it broken down into logical steps. Adjusts method and amount of instruction to the individual patient and is almost always successful.

Executive Ability - Leadership-Organization

S. Always budgets time, plans and executes efficiently. Gives all factors in a situation the correct amount of attention. Adjusts plans easily to changing circumstances. Willing to assume responsibility for making decisions. Directs action of others and is always accepted with confidence.

Initiative

S. Always able to inaugurate plans and operations, has ideas about what should be done in unusual situations and can proceed unassisted. Always "self-starting."

Neatness in Preparation of Work

S. Always has tools and materials readily accessible. Works in an orderly fashion, and always leaves work space in perfect order. Designs are well prepared, measurements and cutting always accurate. Finishing processes precisely and attractively done.

Economy of Materials

S. Knows the value and special qualities of materials in stock and always uses them appropriately with minimum of waste. Always resourceful and discriminating in salvaging and using waste material.

Care of Supplies and Equipment

S. Always keeps track of tools and always puts them away in their proper place. Knows the function of tools and always uses them correctly. Systematically keeps tools and equipment clean and in good working order. Reports damage or need for repairs immediately. Keeps supplies in an orderly, accessible arrangement and protected from deterioration and loss.

Professional Ethics and Etiquette

S. Understands the purpose behind rules and regulations, the necessity for observing channels and customs, and always abides by them. Always courteous and well mannered.

Records - Recognize Importance

S. Understands purposes of records for the hospital, physicians, O. T. department and the student. Reads or practices writing them seriously and to the best advantage. Realizes the importance of concise, clear and trustworthy presentation.

Records - Neatness

S. Records are well typed or legibly written, no obvious corrections or erasures. Always well arranged and professional in appearance.

Records - Accuracy

S. Records are factually correct. Give a true picture. Show proper interpretation and pertinent selection of what has been read, as well as valid observation and precise reporting of patient reaction. Spelling and grammar correct, terminology appropriate. All records kept up to date.

Professional Interest

S. Spares no effort to become successful in O. T. Does extra reading, takes additional courses, voluntarily helps with exhibits, sales and inventories to increase competence. Greatly interested in department organization and coordination of O. T. with other departments. Keeps abreast of current trends in O. T. and allied professions.

Adaptability

S. Always able to adjust to people and situations pleasantly, quickly and efficiently without lowering own standards. Flexible.

Emotional Stability

S. Never reacts unduly to difficult or favorable situations or to the needs of others. Not moody. Keeps professional problems out of personal life and vice versa. Mature.

Poise

S. Gives impression of thorough self-control. Always does the appropriate and graceful thing. Has a well-modulated voice.

Sense of Humor

S. Sees and enjoys the incongruities of life, no matter how subtle. Can laugh at own mistakes and profit by them. Discovers, appreciates and appropriately expresses the ludicrous, but never at the expense of others.

Tact

S. Always sensitive to the needs of others and instinctively does the kindly, gracious thing that makes them comfortable. Can usually make suggestions or give correction without hurting feelings or arousing antagonism.

Judgment

S. Considers all factors in a situation (even in emergencies), weighs them quickly and accurately, comes to logical and wise decisions and knows how to implement them.

Dependability

S. Can be counted on to do the right thing in all situations. Carries pertinent instructions out to the letter. Initiates appropriate action. Never fails to take responsibility.

Intellectual Curiosity

S. Acquires additional knowledge for own satisfaction. Looks up unfamiliar terms. Discusses new situations. Studies departmental records, forms and methods and the reason for their use. Asks intelligent questions at appropriate times and follows up suggested material.

Cooperation with Personnel

S. Always willing to lend a hand and is generous about sharing equipment and facilities. Works unselfishly for group results, subordinating personal likes, dislikes and interests in order to work effectively with others. Always thinks of the effect on others before requesting special consideration.

Ability to Profit by Instruction and Criticism

S. Is a pleasure to instruct. Listens in a receptive and intelligent manner to any help or suggestion offered. Comprehends and utilizes new techniques or concepts. Evaluates criticism. Corrects faults.

Loyalty to Department and Personnel

S. Habitually considers reputation of department and hospital in her speech and actions, on and off duty. Does not gossip. Will give constructive suggestions and criticism to authorized persons only.

Punctuality

S. Always prompt and has all work completed at the appointed time. Does not have to be reminded of schedules or deadlines.

Personal Appearance

S. Observes all regulations pertaining to uniform. Has an erect, graceful carriage, habitually pleasant, alert manner. Neat and well groomed at all times. Creates an extremely favorable professional impression.

Health

S. Has plenty of energy for both work and recreation. Does extra work cheerfully and efficiently. Seldom shows fatigue, rarely ill.

D. Description of Subjects

The group of students used in this research study consisted of all the students in the Department of Physical Medicine of the University of Wisconsin. For this investigation, two hundred and thirty-four Occupational and Physical Therapy students were administered the complete battery of twelve tests. Their ages ranged from 17 years to 45 years of age.

The subjects included 43 Physical Therapy Certificate Course students, 33 Physical Therapy four-year degree course students, and 158 four-year course students in Occupational Therapy. The certificate course for physical therapy technicians is twelve months in length and begins in September of each year. College credit toward a degree in Physical Therapy is given for this work.

Among the forty-three Physical Therapy Certificate Course students were included four male students, and two men students were among the 33 degree trainees in Physical Therapy. Occupational Therapy had no male trainees.

E. Procedures Used in Personnel Research

The subjects were contacted through the mail. The students were sent a postal card informing them of the time, place, and reason for the testing. All of the group tests were administered in Memorial Institute of the University of Wisconsin Hospital. Students unable to meet at the specified time were granted another time. All of the individual tests were given in the Educational Building. All of the tests were given in the evening. Two evenings were necessary for the group tests, and one evening was sufficient time for the tests which were administered individually.

In the administration of the battery of tests, it should be pointed out that the writer was given the sole responsibility for the giving and scoring of the tests. As stated previously, the group tests were given in Memorial Institute and the individual tests were administered in the Education Building. In both situations, the rooms were furnished properly with lighting, equipment, ventilation, and comfortable seats.

One of the most difficult aspects of the administration of the testing program is that which deals with the scoring of tests. Scoring is laborious work, which requires much time and effort on the part of the administrator. This part of the work the present investigator planned with extreme care in order that accurate results may be obtained and in order that the writer may retain favorable attitudes toward the testing program throughout the whole task of scoring. Every precaution was taken to these ends.

Since scoring of the tests was burdensome, every precaution was taken for securing the greatest accuracy with the least amount of time and effort.

The first step in preparing for scoring was for the writer to see to it that he read carefully all the directions accompanying the tests. The second step was to provide a minimum of practice in scoring. The third step was the actual scoring, which was done, as it always should be done, with colored pencils. Because it is most conducive to accuracy and most economical of time the writer scored by single pages of a test, whenever it was applicable.

During the process of scoring, the scheme of checking correct, omitted, and incorrect answers was adhered to without exception. The summary of the scores obtained were placed in appropriate positions on each page and totaled for each test.

After the scoring had been completed, the results of all the tests were placed on the graphic profile of the data form. The test profile lists the names of all the tests, raw scores, grade placements, and percentile ranks so that the position of each pupil may be easily and quickly determined from the graphic profile.

The two-page data form used in this study included the following items: complete name of the student, classification (course and year), sex, race, descent, degrees held, date of birth, marital status, rank in family among brothers and sisters, rank in high school graduating class, number of students in high school graduation class, major course in high school, types of extra-curricular activities.

The data form also included previous wage-earning experiences (kind, employer, and time), reason for choice of course, period of time that the tests were administered, the names of all the tests, raw scores, percentile rank, and psychograph of all the tests.

F. Statistical Procedures Employed

The earliest studies dealing with the prognosis of college success used existing psychological tests. These tests were introduced to predict both college success in general and course achievment in subject areas. At first the primary objective was the prediction of probable college success, but later the emphasis shifted to the prediction (actually a refinement in statistical method) of comparative probability of success in a specific field. Little positive relationship was found between the test results and college achievment in these early studies.

Following this early development and use of intelligence and ability tests at the college level, there was a gradual growth in the areas of personality, aptitude, interest, and judgment testing. It seems that the improvements in the use of these tests have come about through an extension of testing to cover such diverse factors as interest, aptitude, personality, intelligence, achievement, and judgment.

The bare relationship existing between a series of predictive measures and college success measures does not always tell the entire story. Interpretation of test results differs to some extent according to the conditions and circumstances surrounding the individuals being tested.

It was agreed that the purpose of this study was to determine if a given battery of tests could be used to predict success in physical and occupational therapy. The specific questions which needed to be answered were:

1. How well do the predicted grade point average, actual grade point average, and scores on the individual tests correlate with success in therapy?
2. Do the results of Question 1 merit use in predicting success in therapy by a multiple correlation technique? Stated in other words, how much relationship exists between the battery as a whole and success in therapy?
3. How well do the measures correlate with each other? To what extent do they measure the same thing?
4. How well do part scores on the tests correlate with success in therapy?
5. Of what use can these data be in administering guidance to the students?

Prior to approaching the first problem, two subsidiary questions had to be answered: (1) On the tests yielding part scores could the part scores be added together to give a total score, and (2) what criterion was to be used as a measure of success in therapy?

The procedure in which the first question was answered was as follows:

The directions for scoring some of the tests provided methods of computing total scores, either by combination of part scores, or otherwise. For these tests, this method was used. The following instruments were treated in this manner: Wechsler-Bellevue Intelligence Scale, California Mental Maturity, Wide Range Achievement Tests, Michigan Speed of Reading, Purdue Pegboard, Minnesota Paper Form Board, Bennett Mechanical Comprehension, and the Cardall Practical Judgment.

The tests which did not provide total scores and which demanded a different treatment were: Kuder Preference Record, Minnesota Rate of Manipulation, Minnesota Multiphasic Personality Inventory, and the Washburne Social Adjustment Inventory.

In the case of the Kuder Preference Record, it was decided that no total score would suffice. Therefore, two interest scores were chosen, which in the judgment of the committee members and the author of this study, related most highly to success in therapy. These were Scientific-Interest and Artistic-Interest. Scores on these were included as two separate test scores in the first analysis. The other scores were omitted.

In the Rate of Manipulation test, the scores on both parts (placing and turning) were not comparable. The placing test was based on a perfect score of 300, and the turning test on a score of 200. Both scores were treated separately.

Regarding the Multiphasic Personality Inventory, and the Washburne Social Adjustment Inventory it was decided to combine part scores to yield a kind of ''index of adjustment'' resulting from this combination. Data regarding each part score was analyzed.

On the Multiphasic Personality Inventory, a high score on any scale indicates a tendency toward abnormality in that area. This is true on all scales, except the Masculinity-Femininity scale. On this scale a high raw score indicates better adjustment. Therefore, the method used to get the so-called index of adjustment was to sum all part scores except those of Masculinity-Femininity and then subtract from this sum the raw score on Masculinity-Femininity. Thus a high score would indicate roughly a high degree of maladjustment.

On the Washburne, the opposite was the case as a high score in any area indicates a high degree of adjustment in that area. Part scores were simply added to get a total score. Here a high total score indicated a high degree of total adjustment.

The best available estimate of success in therapy was the rating given a student by her supervisor and clinical instructors during her senior year. This subjective rating sheet (exhibit B, Appendix), consisted of a five-point scale constructed by the Department of Physical Medicine. The students were rated on specific traits in three areas. These areas (professional preparedness, personal qualifications, and professional qualifications) were weighted equally. One hundred points were allowed in each area for a total possible score of three hundred. The numerical value of a check mark in any level was indicated on the rating scale.

The total points on each rating scale give the total rating for the student. After transforming all the data into quantitative terms, it was possible to answer Question 1, regarding correlations of test scores, predicted grade point average, and actual grade point average with success in therapy.

The predicted grade point average at the University of Wisconsin is based upon the grade point average earned in high school and the score earned by the student on the American Council on Education Psychological Examination which is given during freshman week. In the case of a transfer student the test is also administered during that week. The formula used for predicting grade point averages is as follows:

$$X_1 \ .0180X_2 + .0197X_3 - 1.1016$$

Formula where X_1 is the predicted grade point average in the University of Wisconsin, X_2 is the American Council on Education Psychological Examination percentile score based on National Norms, X_3 is the high school percentile rank.

Since ratings were available for only those students who had clinical experience, this part of the analysis was limited to 1946 and 1947 graduates. These students totaled only thirty-one in Physical Therapy and forty in Occupational Therapy. Statistically, these did not comprise a sufficient number of cases to warrant the application of the rating scale for obtaining correlations with the test scores, actual grade point average, and the predicted grade point average. The results are shown in columns one and three of Table I.

A broader discussion of these statistical analysis procedures is found in Exhibit A of the Appendix.

SECTION IV

ANALYSIS AND INTERPRETATION OF FINDINGS

A. Analysis of Findings

THE PURPOSE of the statistical analysis was to determine if the battery of the tests chosen, the predicted grade-point average, and the actual grade point average could be used for predicting success in the field of therapy. The specific questions to be answered were:

1. How well do the predicted grade point average, the actual grade point average, and scores on the individual tests correlate with success in therapy?

2. Do the results of Question 1 merit use in predicting success in therapy by a multiple correlation technique? In other words, how much relationship exists between the battery as a whole and success in therapy?

3. How well do the measures correlate with each other? In other words, to what extent do they measure the same thing?

4. How well do part scores on the test correlate with success in therapy?

5. Of what use can these data be in administering guidance to the students?

The mean and percentile scores of the various tests for the entire group of two hundred and thirty-four students are shown in Table I. The means of the Multiphasic Personality Inventory and the Washburne Social Adjustment are not included because, as previously stated, these two instruments did not provide total scores, therefore, they demanded a different treatment. It was decided to combine the part scores of the Multiphasic and Washburne so as to yield a kind of "index of adjustment" resulting from this combination. Data regarding each part score was analyzed.

The Wechsler-Bellevue Intelligence Scale had a mean of 114.901 (80th percentile) which place the score in the above-normal classification; the California MentalMaturity had a mean of 117.233 (85th percentile), and the Wide Range Achievement Test had a mean of 113.259 (79th percentile), both means being in the above-average category.

The means of the Scientific and the Artistic areas of the Kuder Preference Record were 57.642 (67th percentile) and 60.522 (73rd percentile), respectively. These scores indicate above average interests, as any score above the 50th percentile is considered a definite interest pattern.

The mean of the total score of the Purdue Pegboard was 82.034 (83rd percentile); and the mean scores of the Placing and Turning of the Rate of Manipulation Test were 223.720 (74th percentile) and 184.690 (70th percentile), respectively; the Paper Form Board Test had a mean of 44.461 (88th percentile); and the mean scores of the Mechanical Comprehension Test and Practical Judgment were 39.582 (53rd percentile), and 208.065 (73rd percentile).

TABLE I

MEANS AND PERCENTILE SCORES OF VARIOUS TESTS FOR THE ENTIRE GROUP OF STUDENTS*

(N = 234)

Name of Test	Mean	Percentile
Wechsler-Bellevue	114.901	80
California Maturity	117.233	85
Wide Range Achievement	113.259	79
Speed of Reading	57.194	70
Scientific - Kuder	57.642	67
Artistic - Kuder	60.522	73
Purdue Pegboard	82.034	83
Placing - Manipulation	223.720	74
Turning - Manipulation	184.690	70
Paper Form Board	44.461	88
Mechanical Comprehension	39.258	53
Practical Judgment	208.065	73

* The means of the Multiphasic Personality Inventory and the Washburne Social Adjustment Inventory are not included because they are expressed in other values.

The correlations of the test scores, the predicted grade-point average, and the actual grade point average with the ratings on the scale and with the scores obtained from ranking are shown in Table II. * The standard error of these correlations was found to be . 183 for the Physical Therapists and . 160 for the Occupational Therapists. In order to be significant each of these correlations should be twice this large.

The correlations between the ratings of the scale and the rankings were . 43 and . 39 for physical therapy and occupational therapy students, respectively.

The correlations between the predicted grade point average and the physical therapy ratings was . 41 and with the occupational therapy ratings . 36. With the rankings the predicted grade point average correlated . 40 with the thirty-one Physical Therapists and . 41 with the forty Occupational Therapists.

The Turning Test of the manipulation measure correlated . 35 with the original ratings (Physical Therapy) and the Placing part of the manipulation instrument correlated . 31 with the Occupational Therapy group of forty students. The Mechanical Comprehension Test correlated . 32 with the original Occupational Therapy ratings.

The intercorrelations (Table III) which showed a fair amount of relationship were the actual grade point average and the predicted grade point average (. 70); the Wechsler-Bellevue Scale and the predicted grade point average (. 65); the Wechsler-Bellevue and the actual grade point average (.53); the California Mental Maturity and the predicted grade point average (. 53); and the California Mental Maturity with the actual grade point average (. 59); and the Wechsler-Bellevue with the California (. 63).

The Wide Range Achievement Test has intercorrelations of . 51, . 61, . 57, and . 70 with the predicted grade point average, the actual grade point average, the Wechsler-Bellevue Scale, and the California Mental Maturity, respectively. The other achievement test, Michigan Speed of Reading, had correlations of . 46, . 50, . 61, and . 76 with the predicted grade point average, the actual grade point average, the Wechsler-Bellevue, and the Wide Range, respectively.

There was significant positive relationship of the Paper Form Board Test with the following measures: the predicted grade point average (.32); the academic grade point average (. 29); the Wechsler-Bellevue (. 29); the California Mental Maturity (. 35); the Wide Range Achievement (. 20), and the Speed of Reading (. 20).

The Bennet Comprehension had intercorrelations of . 26, . 29, . 28, . 23, . 12, . 19, and . 23 with the predicted grade point average, the academic average, the Wechsler-Bellevue, the California Mental Maturity, the Wide Range, the Speed of Reading, and the Paper Form Board, respectively.

Other significant intercorrelations were the Turning part of the Manipulation test with the Purdue Pegboard (. 40) and with the Placing of the Manipulation instrument (. 56). Also, the Washburne Social Adjustment correlated . 25 with the Multiphasic, . 26 and . 23 with the Placing and Turning of the Manipulation test, respectively.

The Artistic score of the Kuder Preference Record correlated . 27 with the Paper Form Board and . 27 with the Washburne Social Adjustment Inventory.

The standard error of these correlations is . 066, and a correlation to be significant must have an absolute value equal to or greater than . 132, which is twice the standard error.

Statistical consideration was given to find out the extent to which the correlations between the actual grade point average, the predicted grade point average, the instructors' ratings, and the test scores merit use in predicting success in therapy by a multiple correlation technique.

Because of the small number of cases, and also because of a lack of significant correlations, the results do not seem to merit further analysis until more cases are available for study.

An analysis was made concerning the extent to which part scores on the tests correlate with success in therapy as based on the criteria of the predicted grade point average, the actual grade point average, and the instructors' ratings.

The first step was to determine if some of the part scores might be better indices of success in physical therapy than the total scores. The results of this breakdown are shown in Table IV. Here again, only the 31 cases for which ratings were available were used. A correlation to be significant had to have an absolute value of at least . 37. Only a few of the correlations are this large. They are the following: the Verbal score of the California Mental Maturity, . 38; the Musical score of the Kuder Preference Record, . 44; the Truthfulness and Happiness scores of the Washburne Social Adjustment with . 54 and . 40, respectively.

The last phase of the study was concerned with the effectiveness with which the test results, the ratings, the predicted grade point average, the actual grade point average, and the data form

---

*In order to use another method of rating success in therapy, the instructional staff in Occupational Therapy and Physical Therapy ranked each student from one to forty and from one to thirty-one, respectively. These rankings were converted into scores by making them conform to a normal distribution. Correlations were then computed between these converted scores and the test scores, the predicted grade-point average, and the actual grade-point average.

TABLE II

CORRELATIONS OF TEST SCORES WITH RATINGS MADE ON RATING SCALE AND WITH SCORES OBTAINED FROM RANKING

	(N = 31) Physical Therapy Correlation with		(N = 40) Occupational Therapy Correlation with	
Test	Rating	Ranking	Rating	Ranking
Wechsler-Bellevue	.23	.20	-.01	.13
California	.20	.24	.02	.04
Wide Range	.17	.23	.18	.28
Speed of Reading	-.07	.10	-.01	-.07
Scientific - Kuder	-.15	-.24	-.03	-.13
Artistic - Kuder	.13	.26	.01	.12
Purdue Pegboard	.14	.14	.06	-.20
Placing Manipulation	.10	-.08	-.31 x	-.07
Turning Manipulation	-.35 x	-.01	-.22	.01
Form Board	.31	.22	.29	.21
Mechanical Comprehension	.05	-.03	.32 x	.07
Multiphasic	.24	-.14	-.17	.06
Washburne	.31	-.05	-.22	-.15
Judgment	.06	.07	-.07	-.20
Academic average	.28	.30	.22	.14
Predicted grade point average	.41 x	.40 x	.36 x	.41 xx

Standard error = .183
2 × Standard error = .366

Correlation between ranking and rating = .43

x = Significant at 5% level
xx = Significant at 1% level
(.355 and .456)

Standard error = .160
2 × Standard error = .320

Correlation between ranking and rating = .39

x = Significant at 5% level
xx = Significant at 1% level
(.312 and .403)

TABLE III

INTERCORRELATIONS OF TEST SCORES FOR THE TOTAL GROUP
(N = 234)

	P. G. P. A.	Academic Av.	Wech. -Belle.	Calif. M. M.	Wide Range	Speed of Reading	Paper F. Bd.	Mech. Compre.	Multiphasic	Pegboard	Placing	Turning	Washburne	Scientific	Artistic	Judgment
P. G. P. A.																
Academic Av.		.70														
Wech. -Belle.		.65	.53													
Calif. M. M.		.53	.59	.63												
Wide Range		.51	.61	.57	.70											
Speed of Read.		.46	.50	.61	.32	.76										
Paper F. Bd.		.32	.29	.29	.35	.20	.30									
Mech. Compre.		.26	.29	.28	.23	.12	.19	.23								
Multiphasic		-.25	-.25	-.24	-.25	-.14	-.20	-.21	-.16							
Pegboard		.10	.19	.21	.17	.17	.17	.07	.19	-.10						
Placing		-.17	-.17	-.19	-.12	-.00	-.08	.03	-.15	-.03	-.04					
Turning		-.19	.20	-.12	-.19	-.09	.01	-.17	.05	.17	.40	.56				
Washburne		-.15	.11	.06	.04	.12	.04	-.02	.02	.25	.17	.26	.23			
Scientific		.04	.07	.09	.13	.05	.15	.09	.20	-.06	.01	-.34	.01	.04		
Artistic		.03	.03	.08	.01	.15	.09	.27	.02	.09	.07	.07	.00	.27	-.31	
Judgment		.00	.09	.09	.03	.11	.31	.09	.20	-.07	.13	-.01	-.00	.01	.06	-.12

Standard Error = .066
2 × Standard Error = .132

TABLE IV

CORRELATIONS OF PART SCORES WITH RATINGS PHYSIO-THERAPY GROUP (N = 31)

Test	Part	Correlation with Ratings
Wechsler-Bellevue	Verbal	.18
Wechsler-Bellevue	Performance	.26
California	Verbal	.38 x
California	Non-Verbal	.34
Wide Range	Reading	.21
Wide Range	Spelling	.21
Wide Range	Arithmetic	.34
Kuder	Mechanical	.04
Kuder	Computational	-.22
Kuder	Persuasive	.02
Kuder	Literary	.25
Kuder	Musical	.44 x
Kuder	Social Service	.06
Kuder	Clerical	.36
Purdue	Both Hands	.28
Purdue	Assembly	.27
Multiphasic	Question	-.10
Multiphasic	Lie	.03
Multiphasic	Validity	-.19
Multiphasic	Hypochondriasis	-.13
Multiphasic	Hysteria	-.08
Multiphasic	Depression	-.08
Multiphasic	Psychopathic-Deviate	-.18
Multiphasic	Masculinity-Femininity	.07
Multiphasic	Paranoia	-.34
Multiphasic	Psychasthenia	-.11
Multiphasic	Schizophrenia	-.10
Multiphasic	Hypomania	-.16
Washburne	Truthful	.54 xx
Washburne	Happiness	.40 x
Washburne	Alienation	.01
Washburne	Sympathy	.16
Washburne	Purpose	.01
Washburne	Impulse-Judgment	-.00
Washburne	Control	-.19

Standard Error = .183
2 × Standard Error = .366

x = Significant at 5% level
xx = Significant at 1% level

were used by the faculty members of occupational therapy and physical therapy. The Department of Physical Medicine made use of any information or knowledge available from the results of the tests, the data form, and the rating scale.

The Physical Therapy staff used these data in counseling fifty students. These data were used in revealing the abilities, interests, aptitudes, personality traits, personal data, and the limitations of which the students were sometimes unaware.

Five of the fifty physical therapy students, because of poor scholarship, a low predicted grade point average, and through the faculty's use of the test results and rating scale, were advised to drop the course. Indications of these students being unlikely to succeed in the course were evident from the data gathered. The interviewing and advising of all the students was based on the best possible knowledge available about the individual. Three of the students advised to drop the course were eliminated later by the University due to scholastic reasons. The other two students left the course to avoid future disappointment.

The Occupational Therapy staff counseled 80 students. From this number, thirteen trainees were advised to drop the course because of reasons such as low scholarship, unsuitable personality traits, lack of special aptitudes (manual and finger dexterity), conflicting interests, lack of artistic, mechanical, and scientific interests, or for any other significant maladjustment (social, mental, economic, or physical) which caused an inability for one to work to the full extent of his capacity in therapy work. The data used in this guidance phase of the study were gathered from test results, the data form, the interview, and the rating form, all techniques of guidance which seemed to indicate a general prediction that the students would fail as occupational therapy students as well as professional therapists.

Of this group of thirteen, eight students failed to make the scholastic grade; therefore, they were eliminated. The other five, due to changing interests, lack of specific aptitudes, personality problems, lack of judgment, asocial behavior, and low mentality were transferred to another course more suited to their total make-up. In all five cases, guidance was extended the students towards other fields.

It must be emphasized here that the data collected were used in assisting the poor, fair, and good student. Counseling of these students was positive guidance which was developmental and preventive in nature, as well as remedial. Failing students, personality-problem students, change-of-interest students, as well as the better adjusted individuals were counseled.

The effectiveness of these data was also shown when probable and imminent failures and other maladjusted students made good progress after they were counseled and advised in respect to their interests, intelligence, achievement, personality, judgment, predicted level of academic achievement, aptitudes and abilities, and other pertinent factors.

After the students were counseled a case summary of every student was written. These summaries were placed in the Department's files for future reference, along with other findings and information about the students. Summaries of three occupational and physical therapy students are presented in the Appendix, Exhibit N.

These data were used, also, for further consultations when the need had arisen for further guidance. As stated previously, it was the intention of the Department's guidance and testing program to establish positive guidance which was developmental and preventive in nature, as well as remedial. The Department of Physical Medicine wants to install a basic functional guidance program which can be developed so it is an integral part of the total school program of every prospective physical and occupational therapist. The Department's system of studying individuals must be based on the best tested and secured knowledge about the nature of the personality. One cannot understand and counsel individuals without probing into the person's make-up in making an individual inventory.

Following are the criteria which were developed by the committee in conferences on guidance of the therapy students. They are suggested merely as an aid in arriving at some sort of evaluation of a guidance program.

1. Are the needs of the pupil determined?
2. Are the chief interests of each pupil determined?
3. Are the abilities of each pupil determined?
4. Is each pupil's program definitely planned in terms of his interests, his needs, and his abilities?
5. Is each pupil made definitely aware of his special aptitudes or abilities?
6. Is there a provision for systematic and definite collection and recording of data?
7. Does the student understand that the results are to be used to assist him in approaching his problem?
8. Is the number of maladjustments reduced?
9. Is there a placement and follow-up program?
10. Is the number of retarded cases reduced?
11. Is there provision for adequate individual counseling?

B. Interpretation of Findings

There was one total measure which had a sufficiently high positive correlation with the original ratings and the rankings. The correlation which was found to be significant was the predicted grade point average, which correlated .41 with the Physical Therapy ratings and .36 with the Occupational Therapy ratings. Using the rank-order method the correlations were .40 and .41, respectively.

Many of the correlations of the various tests with the ratings were small. This has some significance. It appears that those correlations which are of any size at all may be due to chance. The low correlations with the ratings seem to indicate two possibilities:

1. The rating scale does not give a good evaluation of the student as a therapist, and
2. The committee may have chosen some tests which are not highly related to success in therapy.

It was noted that in cases where more than one rating of the same individual or student was given, the ratings varied widely, that is, different raters gave the same student very different ratings on the same scale.

These results (Table II) also indicate that perhaps the tests used did not measure the same things which were looked for by the others, or raters.

In correlating the measures, as shown in Table III, a fair amount of relationship exists between the actual grade point average and the predicted grade point average (.70); the Wechsler-Bellevue Scale and the predicted grade point average (.65); the Wechsler-Bellevue and the actual grade point average (.53); the California Mental Maturity with the actual grade point average (.59); and the Wechsler-Bellevue with the California (.63).

From all indications, the predicted grade point average, the actual grade point average, and the two intelligence tests, the Wechsler-Bellevue and the California Mental Maturity seem to measure largely the same things and have high correlations.

The achievement tests, the Wide Range and the Speed of Reading, had high intercorrelations with the predicted grade point average, the actual grade point average, the Wechsler-Bellevue, and the California Mental Maturity. It is possible that these two instruments measure many of the same things that the other four measures do, although they do not correlate well with the ratings. This high degree of relationship between the two tests, the Wide Range and the Speed of Reading, and the other four instruments reveals that these measures may be appropriate ones to use to predict success in therapy.

The low intercorrelations of the Mechanical Comprehension Test with all the other scores seem to indicate that the instrument does not measure the same thing as the other tests.

The significant correlation of the Manipulation (rapid and accurate finger and hand movements) with the Pegboard (tip-of-the-finger movements and finer movements) warrants further use of the instrument in future research.

There appears to be little or no relationship between the personality and the social adjustment tests and the interest inventory with the rest of the tests. From Table III, it is evident that the instruments seem to measure fairly unique, fine, and amenable traits of the individual. These three measures may or may not be appropriate ones from which to predict success in therapy.

Of the four part scores which might be better indices of success in physical therapy than the total scores only two seem to warrant further study with more cases, providing the rating scale does not give a good evaluation of the student as a therapist. These are the Happiness score and the Truthfulness score on the Washburne Inventory. There was not a sufficient number of cases available to correlate part scores on the tests with the ratings. Only 31 cases were available for use, therefore, the low correlations of all the part scores with the ratings may be due to one of three things:

1. Not enough cases were available.
2. The rating scale may not give a good evaluation of the student as a therapist.
3. The tests may not be appropriate ones from which to predict success in therapy.

Because of the small number of cases, and, also, because of a lack of positive significant correlations, the results did not seem to merit further analysis in predicting success in therapy by a multiple correlation technique until more cases were available.

The value of the guidance program in the Department of Physical Medicine is almost wholly dependent upon the effectiveness with which the results are used by the faculty. Intelligent and wide use of the data (test results, ratings, and other pertinent information) calls for a definite plan of testing and of organizing and reporting the results and rendering them understandable to the students and faculty.

The Department has adopted a policy and a systematic program using these data to counsel students in connection with problems involving diagnosis, prognosis, treatment, and counseling. Helpful suggestions were extended to all of the students. Students not suited for occupational therapy or physical therapy were advised to change to a course more suited to their interests, aptitudes, intelligence, personality traits, and other factors. Other students, who appeared likely to succeed in therapy were advised and counseled as to their progress. Failing students, as well as the more adjusted trainees, were given thorough guidance and help in attaining possible objectives.

The following reasons were given for the causes of failures, drop-outs, and maladjustments:

1. Low scholarship
2. Low intelligence or mental ability
3. Change of interest and conflict of interests
4. Personality deviations
5. Lack of specific aptitudes (finger and manual dexterity)
6. Asocial and antisocial behavior
7. Low actual and predicted level of achievement
8. Maladjustments (emotional, social economic, mental, or physical)
9. Poor mental or physical health

The value of this guidance program was noted by:

1. The number of cases in which the cause of maladjustment was eliminated and the individual student was enabled to work to the full extent of her capacity.
2. The number of cases in which improvement was shown in making the proper adjustment.
3. The number of cases in which the corrective adjustment seemed to have been permanent.
4. The number of drop-outs prevented.
5. The number of failures prevented.
6. Good adjustment made by students in other fields who left Occupational and Physical Therapy due to change of interest, personality problems, lack of certain aptitudes, or for any other significant maladjustment, who received valuable assistance and counseling from the faculty of the Department of Physical Medicine.

SECTION V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

An attempt has been made to develop a program which might provide reliable information for judging the suitability of a student for physical therapy or occupational therapy work. Twelve tests were administered experimentally and the results of these tests were compared with the subjective rating of the students by their instructors and clinical supervisors during their training period. The committee with this study established the following primary objectives in the program:

1. To provide accurate and meaningful measurements of the individual's interests, intelligence, aptitudes, personality traits, achievement, mechanical aptitudes, and judgment.

2. To assist in the selection of students with potentialities to make good physical or occupational therapy students.

3. What is a reasonable expectation of achievement in physical or occupational therapy by each individual?

4. To what extent is each person limited by intelligence, interest, aptitude, judgment, and personality?

5. Is success related to intelligence, personality, interest, aptitudes, achievement, and judgment?

This study consisted of two parts. The first part involved the administration of twelve tests to 234 students of physical therapy and occupational therapy.

The second part of the study was concerned with the rating of the students with a five-point rating scale which was concerned with professional preparedness, personal qualifications, and professional qualifications.

The purpose of the statistical analysis was to determine if the battery of tests chosen, the predicted grade point average, and the actual grade point average could be used for predicting success in the field of therapy. The specific questions to be answered were:

1. How well do the predicted grade point average, the actual grade point average, and the scores on the individual tests correlate with success in therapy?

2. Do the results of Question 1 merit use in predicting success in therapy by a multiple correlation technique? In other words, how much relationship exists between the battery as a whole and success in therapy?

3. How well do the measures correlate with each other? In other words, to what extent do they measure the same thing?

4. How well do part scores on the test correlate with success in therapy?

5. Of what use can these data be in administering guidance to the students?

A careful survey of available information from published reports and references was made and information was gathered from the University of Wisconsin Library and the Department of Physical Medicine. The survey of literature also included books and pamphlets listed in the Education Index and the Reader's Guide to Periodical Literature.

The data presented in this study were secured at the University of Wisconsin by using a battery of tests, a rating scale, a data form, and other information gathered from the subjects, the Student Counseling Center, and the Registrar's office.

The predicted grade point average was the only total measure which had a very high significant correlation with the original ratings and rankings. It correlated .41 and .36 with the Physical Therapy and Occupational Therapy students' ratings, respectively, and using the rank-order method the correlations were .40 and .41, respectively. The correlations between the ratings of the scale and the rankings were .43 and .39 for Physical Therapy and Occupational Therapy students, respectively.

Many of the correlations of the various tests with the ratings were small. It appears that those correlations which are of any size at all may be due to chance. The low correlations with the ratings seem to indicate two possibilities:

1. The rating scale does not give a good evaluation of the student as a therapist, and

2. The committee may have chosen some tests which are not highly related to success in therapy.

A fair amount of relationship exists between the actual grade point average and the predicted grade point average (.70); the Wechsler-Bellevue Intelligence Scale and the predicted grade point average (.65); the Wechsler-Bellevue and the actual grade point average (.53); the California Mental Maturity with the actual grade point average (.59); and the Wechsler-Bellevue with the California (.63).

From all indications, these instruments seem to measure the same things and have high correlations which warrant further use in future research and study with success in therapy.

The achievement tests, the Wide Range and the Speed of Reading, had high correlations and intercorrelations with the predicted grade point average, the actual grade point average, the Wechsler-Bellevue, and the Mental Maturity. These measures seem to be appropriate ones to use to predict success in therapy.

There was positive relationship of the Paper Form Board Test with the following measures: the predicted grade-point average (.32); the actual grade point average (.29); the Wechsler-Bellevue (.29); the Mental Maturity (.35); the Wide Range Achievement (.20); and the Speed of Reading (.30). This instrument does not seem to measure the same things that most of the other instruments do.

The Comprehension Test had intercorrelations of .26, .23, .12, .19, and .23 with the predicted grade point average, the academic average, the Wechsler-Bellevue, the California Mental Maturity, the Wide Range Achievement, the Speed of Reading, and the Paper Form Board, respectively. Though these correlations and intercorrelations were low with other measures, this small relationship has nothing to do with the value of the test in predicting success in Physical Medicine. It may be the best instrument of the entire battery. If the Mechanical Comprehension Test does not correlate with success in therapy then it should be discarded.

The Turning part of the Manipulation Test correlated .40 with the Purdue Pegboard and .56 with the Placing of the Manipulation Test. This test, to some extent, seems to be an appropriate one from which to measure success in therapy.

The Washburne Social Adjustment Inventory correlated .25 with the Multiphasic Personality Inventory; .26 and .23 with the Placing and Turning of the Manipulation Test, respectively. There appears to be little relationship between these two personality measures with the rest of the tests. It should be clearly stated that low intercorrelations with other measures has nothing to do with the value of the test in predicting success in therapy.

The next approach was concerned with the extent to which the correlations between the predicted grade point average, the actual grade point average, the instructors' ratings and the test scores merit use in predicting success in therapy by a multiple correlation technique. Only 31 cases were available. Because of the small number of cases, and also, because of a lack of significant correlations, the results did not seem to merit further analysis in predicting success in therapy by a multiple correlation technique until more cases were available.

It is evident that the instruments seem to measure fairly unique traits of the individual, but these measures may or may not be appropriate ones from which to predict success in therapy.

The Artistic score of the Kuder Preference Record correlated .27 with the Paper Form Board and .27 with the Washburne Social Adjustment Inventory. This test may be an appropriate one from which to predict success in therapy.

The fourth question was concerned with the extent to which the part scores of the tests correlate with success in therapy as based on the criteria of the predicted grade point average, the actual grade point average, and the instructors' ratings.

The first step was to determine if some of the part scores might be better indices of success in physical therapy than the total scores. Only the 31 cases for which ratings were available were used. A correlation to be significant here had to have an absolute value of at least .37. Only a few of the correlations were this large.

The part scores and their correlations follow: the Verbal score of the Mental Maturity Test .38; the Musical score of the Kuder Preference Record .44; the Truthfulness and Happiness scores of the Washburne Social Adjustment Inventory .54 and .40, respectively.

Of the four which were statistically significant two of them seemed to warrant further study with more cases, without any questions asked. These are the Happiness score and the Truthfulness score on the Washburne Inventory. Statistically speaking, there was not a sufficient number of cases available to correlate part scores on the tests with the ratings. The low correlations of all the part scores with the ratings may be due to one of three things:

1. Not enough cases were available.
2. The rating scale does not give a good evaluation of the student as a therapist.
3. The tests may not be appropriate ones from which to predict success in therapy.

The last phase of the study was concerned with the effectiveness with which the test results, the ratings, the predicted grade point average, the actual grade point average, and the data form were used by the faculty members of occupational and physical therapy.

The Physical Therapy staff used these data in counseling fifty students, and the Occupational Therapy staff advised eighty students. The students were counseled concerning their interests, aptitudes, personality traits, personal data, and the limitations of which the students were sometimes unaware. It must be emphasized that the data were used in assisting the poor, fair, and good student.

Five of the fifty physical therapy students, because of some good reason, were advised to drop the course. Thirteen trainees in Occupational Therapy were advised to drop the program. After the students were counseled, a case summary of every student was written.

The effectiveness of these data was also shown when imminent and probable failures made good progress after they were counseled.

Conclusions

From the analysis of the data collected, the following conclusions seem warranted:

1. The rating scale does not contain the most important traits essential for success in therapy, and the right things are not being judged. There is too much emphasis on the individual's personality, and not enough factors which are more concerned with the success of the student in therapy.

2. The various rated factors are not weighed properly in the total evaluation.

3. The raters are not able to discriminate enough, perhaps because of insufficient contact with the student, or because the traits rated are not defined adequately, or because of inability to actually recognize traits, or lastly, because of the influence of personal bias, the halo effect, and the error of lenience or severity. There may have been some hesitation to give ratings at the extremes of the scale.

4. Some of the traits in the scale are not amenable to rating.

5. There was only one total measure which had a significant correlation with the ratings. This was the predicted grade point average.

6. The low correlations of the tests with the ratings and the rankings seem to indicate two possible things:

   a. The rating scale does not give a good evaluation of the student as a therapist, and

   b. Some of the tests chosen may not be greatly or highly related to success in therapy.

7. From all indications, the predicted grade point average, the actual grade point average, the two intelligence tests (Wechsler-Bellevue and the California Mental Maturity) and the achievement tests (the Wide Range Achievement and the Speed of Reading) seem to measure the same things and have high correlations. Through more study one may find these instruments to be appropriate ones to use to predict success in therapy. The claim cannot be made here that these measures should be used in future research because of the high correlations.

8. The Kuder Preference Record and the personality tests (Multiphasic Personality and the Social Adjustment Inventory) measure

many amenable and fine traits of the individual. There appears to be little relationship between these tests and the rest of the battery. Again, the claim should be made here that the low correlations and intercorrelations of these instruments with other measures has nothing to do with the value of the tests in predicting success in therapy.

9. The Manipulation Test correlates fairly well with the other measures and with the ratings, which substantiates the statement and conclusion that the instrument should be used in future research from which to predict success in therapy.

10. The Pegboard Test, the Paper Form Board, the Mechanical Comprehension scale, and the Practical Judgment instrument showed little or no relationship with the ratings, or with the rest of the tests. These instruments may be or may not be appropriate ones from which to predict success in therapy. Since weaknesses of the rating scale have been established it is proper to reserve for future research the status of these four instruments in the present battery.

11. Only two part scores showed any relationship with the ratings. These were the Happiness score and the Truthfulness score on the Washburne Social Adjustment Inventory. The low correlations of all the part scores with the ratings may be due to one of three things:

    a. Not enough cases were available.

    b. The tests may not be appropriate ones from which to predict success in therapy.

    c. The rating scale does not give a good evaluation of the student as a therapist.

12. Not enough cases were available for having correlations of the part scores of the tests with the instructor's ratings, therefore, the results did not seem to merit further analysis with the rating scale.

13. Because of the small number of cases, and, also because of a lack of significant correlations, the multiple correlation technique was not useful.

14. The value of the guidance program in the Department of Physical Medicine was almost wholly dependent upon the effectiveness with which the results of the tests, the ratings, and pertinent information from the data form, were used by the faculty and its staff. The data were used wisely and properly by the Department in counseling the good, fair, and poor student and trainee according to her interest, intelligence, aptitudes, personality, judgment, and level of achievement or attainment.

15. The Department of Physical Medicine realized that if it is to do an effective job of selection and guidance of its students, the faculty must be provided with dependable, meaningful, objective and subjective, and accurate information about each individual. The faculty members realized they must be thoroughly informed concerning the meaning and uses of this information.

16. Unless the criteria of success are clearly designated the evidence is quite meaningless.

17. This study has been concerned with the subjects for too brief a period (from three months to two years). Unless information and evidence accrues from follow-up studies over a span of years, the prognostic value of the measures used in this research is likely to remain unknown.

18. To do justice to the battery of tests used in this study more study and another experiment should be made with the tests in using the instruments with a refined, revised, and improved rating scale which measures success in therapy.

19. No test of this battery or any combination of tests in this battery measured all of the objectives set out for in measuring appropriate traits for success in therapy. Nor did it weight those it did measure the same as those concerned. This was clearly understood and accepted because no instrument is perfectly valid even for the purpose for which it is designed.

20. The low intercorrelations of the instruments with other measures has nothing to do with the value of the test in predicting success in Occupational Therapy and Physical Therapy. It may be the best measure of the battery. If it does not correlate with success in therapy then it should be discarded.

Recommendations

The foregoing findings prompt the author to recommend certain matters for further investigation and study. Consideration should be given to the following problems:

1. A job analysis and a job specification of both a physical therapist and an occupational therapist should be made by the faculty members of the Department of Physical Medicine, so the raters will know what is expected of a good therapist.
2. The present rating scale should be revised and improved so that every factor essential to success in therapy is included. Every factor should have its proper weighting.
3. The following measures should be given more investigation and experimentation before being replaced by instruments which will measure the same things as the other ten measures: The Mechanical Comprehension Test, the Practical Judgment, the Paper Form Board, and the Purdue Pegboard.
4. The following ten instruments should be used in future research because they measure to some degree success in therapy: The Wechsler-Bellevue Intelligence Scale, the California Mental Maturity, the Wide Range Achievement, the Speed of Reading, the Multiphasic Personality Inventory, the Washburne Social Adjustment Inventory, the Rate of Manipulation Test, the Kuder Preference Record, the predicted grade point average, and the actual grade point average.
5. Since there was little correlation between the tests and the rating scale it is necessary to change the rating scale, or revise the present battery of tests, or develop a group of instruments which will correlate.
6. Some parts of certain tests seem to be more accurate in recognizing the successful students than others. This should be investigated more fully with a larger sampling of trainees.
7. The Department of Physical Medicine should develop as soon as possible tests on the refinement of motor ability and patient-therapist relationship. Considerable attention should be given to the refinement of available personality measurements in approaching the problem of a patient-therapist test.
8. This experimental study should be continued until sufficient criteria are established for a successful program in selecting students for Physical Medicine.
9. Since selective devices are necessary for the selection of students in Physical Medicine, this program should be adopted by other schools in the country in selecting their physical therapy and occupational therapy students.
10. A cumulative record should be constructed by the Department of Physical Medicine so that it may include in composite and meaningful form the data received by the Department on all of its students from the test results, interview, data form, and other effective techniques of guidance.
11. The Department of Physical Medicine should arrange with every student of therapy a pretest interview and a post-test interview. Under such an arrangement the students may discuss their whole program prior to and after an appraisal and an evaluation has been made. Negative guidance, drop-outs, and maladjusted cases can be decreased if the students have this faculty-student relationship.

Needed Research

Several studies need to be made in this experimental and developmental area of the selection of students for the Department of Physical Medicine. As research in prediction and selection for college and vocational success now stands there is a definite need for research study in this experimental study with:

1. More valid and reliable measures of factors that seem to play a large part in the success of the individual in physical therapy and occupational therapy.
2. Studies for the factors and of the factors that enter into satisfactory occupational adjustment as well as educational, personal, and social adjustment of the occupational therapist and physical therapist.
3. A comprehensive follow-up study over a period of years of the criteria used in the selection of students in Physical Medicine.
4. Further investigation required to ascertain the significance of the numerous characteristics rated to the therapy student's total adjustment.
5. A job analysis, job specification, a job description, and an instructional analysis of the occupational therapist and physical therapist by the faculty members of the Department of Physical Medicine. The faculty members know what is expected of a good therapist.
6. A reconstruction or revision of the present rating scale so that it includes all the factors that make for success in therapy.

BIBLIOGRAPHY

1. Adams, H. F. and others. ''Personality as Revealed by Mental Test Scores and School Grades,'' Journal of Applied Psychology, XII (1928), pp. 261-77.

2. Archer, Virginia. A Study of the Social Adjustment of College Students (Syracuse, New York: Syracuse University Press, 1938), 239 pp.

3. Balinsky, B., and others. ''Relative Effectiveness of the Stanford-Binet Scale, etc., and the Wechsler Scale,'' American Journal of Orthopsychiatry, IX (1939), pp. 798-801.

4. Barr, A. S. ''The Coefficient of Correlation,'' Journal of Educational Psychology, XXIII (January 1931), pp. 55-60.

5. Bear, R. M. ''The Predictive Value of the Iowa Physics Aptitude Placement Test,'' Journal of Applied Psychology, XI (October 1927), pp. 381-84.

6. Billett, Roy O. Provision for Individual Differences, Educational Monograph (Washington, D.C.: U. S. Government Printing Office, 1933), pp. 424-72.

7. Bingham, W. V. Aptitudes and Aptitude Testing (New York: Harper and Brothers, 1937), 456 pp.

8. Blair, J. L. Significant Factors in the Prediction of the Success of College Freshmen (Chicago: University of Chicago Press, 1933) pp. 67-145.

9. Bradshaw, F. F. ''The American Council on Education Rating Scale,'' Archives of Psychology, No. 119 (1930), pp. 74-76.

10. Bridges, J. W. ''The Correlation Between College Grades and the Army Alpha Intelligence Tests,'' Journal of Educational Psychology, XI (1920), pp. 361-67.

11. Buros, Oscar K. Mental Measurements Yearbook (New Brunswick, N. J.: Rutgers University Press, 1941).

12. California Test of Mental Maturity Manual (Los Angeles: California Test Bureau, 1946).

13. Cardall, Alfred J. Manual on Practical Judgment Test (Chicago: Science Research Associates, 1942).

14. Clark, Harold. ''Occupational Choices of Today's Youth,'' Vocational Guidance Magazine, Occupations, (February 1934), pp. 34-38.

15. Clark, Harold. ''Occupational Trends,'' Vocational Guidance Magazine, Occupations (February 1934), pp. 46-48.

16. Conklin, E. S. and Sutherland, J. W. ''A Comparison of the Scale of Values Method with Order-of-Merit Method,'' Journal of Experimental Psychology, VI (1923), pp. 44-46.

17. Conrad, H. S. ''The Bogey of the Personal Equation in Ratings of Intelligence,'' Journal of Educational Psychology, XXIII (1923), pp. 147-49.

18. Crane, E. ''The Results of Some Psychological Tests at Bryn Mawr College,'' School and Society, XXV (May 28, 1927), pp. 640-44.

19. Crawford, J. E. Measurement of Some Factors Upon Which is Based Achievement in Elementary Drafting (Pittsburgh: University of Pittsburgh Press, 1941), 125 pp.

20. Cuff, Noel B. ''Prognosis and Diagnosis of Success in College,'' Journal of Applied Psychology, XIV (December 1930), pp. 612-19.

21. Darley, John G. ''A Study of Clinical Predictions of student Success or Failure in Professional Training,'' Journal of Educational Psychology, XXIX (1938), pp. 335-354.

22. Darley, John G. Testing and Counseling in the High School Today, Guidance Program (Chicago: Science Research Associates, 1945), 222 pp.

23. Dickinson, A. ''A Study of Professional Schools—Their Admission and Selection,'' Journal of Engineering Education, XXIX (1939), pp. 385-87.

24. Douglass, Harl R. The Relation of Vocational Success and Certain Factors of Academic Success at the University of Minnesota (Minneapolis: University of Minnesota Press, 1941), 69 pp.

25. Douglass, H. R. and McCullough, C. M. Prediction of Success in School of Dentistry (Minneapolis: University of Minnesota, 1942), pp. 61-74.

26. Fisj, Jean M. Character Ratings and Aca-

demic Grades (Madison, Wis.: University of Wisconsin Press, 1929), 27 pp.

27. Freeman, Frank S. "Predicting Academic Survival," Journal of Educational Research, XXIII (February 1931), pp. 113-23.

28. Freeman, Frank S. "Probable College Success," Journal of Educational Research XXXIV (May 1938), pp. 45-56.

29. Freyd, Max. "The Graphic Rating Scale," Journal of Educational Psychology, XIV (1923), pp. 83-102.

30. Freyd, Max. "Comments on the Graphic Rating Scale," Journal of Educational Psychology, XIV (1923), 94-95.

31. Fry, Dinah E. Graduate Records Bureau (New York: Foundation for Advancement of Teaching, 1945).

32. Fryer, Douglas. "Occupational Intelligence Standards," School and Society, XXV (1922), pp. 408-14.

33. Furfey, P. H. "An Improved Rating Scale Technique," Journal of Educational Psychology, XVII (1926), pp. 45-48.

34. Garrett, H. E. Statistics in Psychology and Education (New York: Longmans, Green and Co., 1926).

35. Gaudet, F. J. and Marryatt, F. J. "Predictive Value of the Law Aptitude Examination," American Law School Review, VII (1930), pp. 27-32.

36. Greene, H. A. and others. Measurement and Evaluation in the Secondary School (New York: Longmans, Green and Co., 1944).

37. Gregg, Alann. "The Future of Physical Medicine," Journal of American Medical Association, CXXV (August 19, 1944), pp. 1093-97.

38. Gregg, Alann. "Trends in Physical Therapy," Journal of American Medical Association, CXXIII (June 4, 1943), pp. 657-703.

39. Guilford, J. P. "A Prognostic Test for Students in Design," Journal of Applied Psychology, XXV, pp. 335-397.

40. Guilford, J. P. Psychometric Methods (New York: McGraw-Hill Book Co., 1936), 566 pp.

41. Hahn, M. E. and Williams, CT. "The Measured Interests of Marine Corps Women Reservists," Journal of Applied Psychology, XXIX (1945), pp. 198-211.

42. Hammond, H. P. and Stoddard G. D. Iowa Placement Examination Study (Iowa City: University of Iowa Press, 1926), 213 pp.

43. Harris, A. J. "The Relative Significance of Measures of Mechanical Aptitude, Intelligence, and Previous Scholarship for Predicting Achievement in Dental School," Journal of Applied Psychology, XXI (1937), pp. 513-521.

44. Hathaway, S. R. and McKinly, J. C. "A Multiphasic Personality Schedule," Journal of Clinical Psychology, X (1940), pp. 249-254.

45. Hathaway, S. R. "Personality Inventory as Aid in the Diagnosis of Psychopathic Inferiors," Journal of Consulting Psychology, III, pp. 112-17.

46. Hull, C. L. Aptitude Testing (Yonkers, N. Y.: World Book Co., 1928), 315 pp.

47. Jastak, Joseph. Wide Range Achievement Test (New York: Charles L. Story Co., 1946), 24 pp.

48. Jollingworth, H. L. "Judging Human Character," Century Magazine, (1922), pp. 45-56.

49. Jurgenson, Clifford E. "Extension of Rate of Manipulation," Journal of Applied Psychology, XXVII (April 1943), pp. 164-69.

50. Kandel, Isaac. Professional Aptitude Tests (New York: Bureau of Publications, Teachers College, Columbia University, 1940).

51. Kelley, Truman L. Statistical Method (New York: Macmillan and Co., 1924), 356 pp.

52. Kelley, Truman L. Tests and Measurements in the Social Studies (New York: Scribner Publishers), 635 pp.

53. Kelley, Truman L. The Stanford Achievement Test (Yonkers, N. Y.: World Book Co., 1938), 63 pp.

54. Knower, F. H. "Studies of Tests in Attitudes," Journal of Social Psychology, VI pp. 315-47

55. Kuder, Frederic G. Kuder Preference Record (Chicago: Science Research Associates,

56. Kuhlmann, Frank. “Binet System for Measuring Intelligence,” *Journal of Psycho-Asthenics,* XV, pp. 76-92.

57. Lane, H. S. “Relationship of Mental Tests to Educational Achievement,” *Volta Review,* XLVIII (November 1946), p. 734.

58. Lindquist, E. F. *Statistical Analysis in Educational Research* (New York: American Book Co., 1940).

59. Lehman, Ruth T. “Interpretation of the Kuder Preference Record for College Students of Home Economics,” *Educational and Psychological Measurement,* III (1943), pp. 217-24.

60. Likert, R. and Quasha, W. H. *Revised Minnesota Paper Form Board* (New York: New York University Press), 367 pp.

61. McCullough, C. M. *Prediction of Success in School of Dentistry at University of Minnesota* (Minneapolis: University of Minnesota Press, 1938).

62. McKinney, Fred. *Psychology of Personal Adjustment* (New York: John Wiley and Sons, 1941), 237 pp.

63. McKinley, J. C. and Hathaway, S. R. “A Multiphasic Personality Schedule,” *Journal of Psychology,* X (1940), pp. 255-68.

64. McKinley, J. C. and Hathaway, S. R. “The Identification and Measurement of the Psychoneuroses in Medical Practice,” *Journal of the American Medical Association,* CXXII (1943), pp. 161-67.

65. MacPhail, A. H. and Joslin, H. L. “Psychological Testing in Schools of Nursing,” *American Journal of Nursing,* XXXII (August 1932), pp. 875-84.

66. Mangold, B. J. *An Analysis of the Kuder Preference Record,* M. A. Thesis, MacMurray College, Jacksonville, Illinois, 1944, 417 pp.

67. May, Mark, and Hartshorne, Hugh. *Studies in Deceit* (New York: Macmillan and Co., 1928), pp. 414-417.

68. McCall, W. A. and Hoerster, L. M. *How to Measure in Education* (New York: Macmillan Co., 1922), 172 pp.

69. Miner, J. B. “The Evaluation of a Method for Finely Graduated Estimates of Abilities,” *Journal of Applied Psychology,* I (June 1937), pp. 123-133.

70. Monroe, W. S. *Encyclopedia of Educational Research* (New York: Macmillan Co., 1941), 888 pp.

71. Moss, F. A. “Report of the Committee on Aptitude Tests for Medical Students,” *Journal of the Association of American Medical Colleges,* X (January 1935), pp. 23-45.

72. Newcomb, Theodore. “An Experiment Designed to Test the Validity of a Rating Technique,” *Journal of Educational Psychology,* XXII (1931), pp. 279-89.

73. *Occupational Therapy: Principles and Practices of Rehabilitation* (New York: A. S. Barnes and Co., 1943).

74. *Occupational Therapy* (Chicago: Science Research Associates, 1944), pp. 47-49.

75. “Occupational Therapy Manual,” *Physical Therapy and Occupational Therapy Magazine* (1943).

76. “Occupational Therapy,” *Occupational Therapy Monograph* (1947).

77. “Occupational Therapy,” *Occupational Therapy Monograph* (1947).

78. Odell, Charles W. “An Attempt at Predicting Success in the Freshman Year at College,” *School and Society,* XXV (June 11, 1927), pp. 702-706.

79. Parker, Marion K. *Rating Scales in Education,* M. A. Thesis, University of Colorado, 1933.

80. Paterson, D. G. “Methods of Rating Human Qual,” *Annals of American Academy Of Political and Social Science,* CX (November 1923), pp. 81-93.

81. *Physical Therapy Handbook,* 3rd Edition (Chicago: American Medical Association, 1939).

82. Pressey, S. L. “An Attempt to Measure the Comparative Importance of General Intelligence and Certain Character Traits in Contributing to Success in School,” *Elementary School Journal,* XXI (1927), pp. 220-234.

83. Purdue Research Foundation. *Manual on Purdue Pegboard* (Chicago: Science Research

Associates, 1942).

84. Rabin, F. G. "Early Conceptions of Measuring Instruments," *Journal of Experimental Psychology*, II (June 1936), pp. 178-224.

85. Remmers, H. H. and others. "Attitude Scales," *Journal of Social Psychology*, V, pp. 298-312.

86. Roberts, A. C. "Objective Measures of Intelligence in Relation to High School and College Administration," *Educational Supervision and Administration*, VIII (1922), pp. 530-40.

87. Richards, C. A. *Modern Clinical Psychology* (New York: Scribner and Sons, 1945), 289 pp.

88. Ruch, G. M. "College Qualifying Examinations," *School and Society*, XXI (1925), pp. 583-86.

89. Rugg, Harold O. "Is the Rating of Human Character Practicable?" *Journal of Educational Psychology*, XII (1922), pp. 38-40.

90. Sackett, R. L. "Engineers Council for Professional Development," *Journal of Engineering Education*, XXXVIII (1937), pp. 502-503.

91. Sangren, P. V. "Social Rating of Best and Poorest High School Students, 289 Traits," *Journal of Educational Psychology*, XIV (1923), pp. 209-14.

92. Schultz, Richard. "The Relation of General Intelligence, Motor Adaptability, and Motor Learning to Success in Dental Courses," *Journal of Clinical Psychology*, XXI (1921), pp. 226-34

93. Segel, David. "Differential Prediction of Scholastic Success," *School and Society*, XXXIX (January 20, 1934), pp. 91-96.

94. Segel, David. "Prognosis of Singing Ability," *Journal of Applied Psychology*, XIII, pp. 425-439.

95. Segel, David. "Comments on Achievement Examinations," *Journal of Educational Research*, XXVIII (1938), pp. 232-36.

96. Shen, Robert. "The Reliability Coefficient of Personal Ratings," *Journal of Educational Research*, XVI (1925), pp. 232-36.

97. Segel, David. *Prediction of Success in College*, Bulletin No. 15 (Washington, D. C.: U. S. Office of Education, 1934), 98 pp.

98. Segel, David. "Tests and Measurements in High School Instruction," *Journal of Educational and Psychological Measurement*, XIX (1930), pp. 178-87.

99. Snedden, D. S. *Study in Disguised Intelligence Tests* (New York: Columbia University Press, 1937), 48 pp.

100. Spearman, C. "Pitfalls in the Use of Probable Errors," *Journal of Educational Psychology*, XXIV, pp. 94-97.

101. Stagner, Ross. "The Relation of Personality to Academic Aptitude and Achievement," *Journal of Educational Research*, XXVI (1933), pp. 648-68.

102. Strang, Ruth. *An Evaluation of Reading Tests for College Students* (1941), pp. 35-37.

103. Strang, Ruth. *Counseling Technics in Colleges and Secondary Schools* (New York: Harper and Brothers, 1937), 159 pp.

104. Strong, Edward K. "The Interests of Social Service Men," *Journal of Educational Psychological Measurement*, V (1945), pp. 157-72.

105. Sullivan, Helen Blair. *Capacity and Achievement Tests* (Yonkers, N. Y.: World Book Co., 1938), 29 pp.

106. Symonds, Percival M. *Diagnosing Personality and Conduct* (New York: D-Appleton-Century Co., 1931), 75 pp.

107. Symonds, Percival M. "On the Loss of Reliability in Ratings Due to Coarsness of the Scale," *Journal of Experimental Psychology*, VII, pp. 456-61.

108. Traxler, A. E. "Some Data on the Kuder Preference Record," *Journal of Educational and Psychological Measurement*, I (July 1941), pp. 253-68.

109. Terman, L. M. and Miles, C. C. *Sex and Personality* (New York: McGraw-Hill Book Co., 1936), 600 pp.

110. Thomson, G. H. *The Factorial Analysis of Human Ability* (New York: Houghton Mifflin Co., 1939), 256 pp.

111. Thorndike, E. L. *Prediction of Vocational Success* (Oxford, England: Oxford Univer-

sity Press, 1934), 224 pp.

112. Thurstone, L. L. The Theory of Multiple Factors (Chicago: University of Chicago Press, 1935), 213 pp.

113. Traxler, Arthur E. "Review of the Kuder Preference Record," Mental Measurements Yearbook (1941), pp. 449-51.

114. Travis, Lee. Educational Diagnosis, 34th Yearbook, National Society for Study of Education (Bloomington, Ill.: Public School Publishing Co., 1935), 563 pp.

115. Triggs, Frances O. "A Study of the Relation of the Kuder Preference Records to Various Other Measures," Journal of Educational and Psychological Measurement, III (1943), pp. 341-54.

116. Veterans Administration Manual, Occupational Outlook, (August 1946).

117. Walker, Helen. Elementary Statistics (New York: Harper and Brothers, 1939), 199 pp.

118. Washburne, John N. "Washburne Social Adjustment Inventory," Mental Measurements Yearbook (New Brunswick, N. J.: Rutgers University Press, 1939), pp. 326-331.

119. Webb, E. "Character and Intelligence," British Journal of Psychology, Monograph Supplement, VII (January 1939), pp. 109-23.

120. Wechsler, David. Measurement of Adult Intelligence Scale (Baltimore: Williams and Wilkins Co., 1941), 229 pp.

121. Wells, F. L. "A Statistical Study of Literary Merit," Archives of Psychology, No. 7 (1907), pp. 45-85.

122. Wightwick, M. Irene. Vocational Interest Patterns (New York: Bureau of Publications, Teachers College, Columbia University, 1945), 145 pp.

123. Wilson, M. D. and Hodges, J. H. "Predicting Success in Engineering College," Journal of Applied Psychology, XVI (1932), pp. 343-357.

124. Wolfson, C. S. "Correlations of Social and Intellectual Forces," School Review, XLII, pp. 53-58.

125. Yum, L. A. "The Stability of Occupational or Preference Items," Journal of Social Psychology, X (April 1939), pp. 41-50.

APPENDICES

* *

Appendices D through M (copies of the tests) are omitted here and may be found in the original study on file in the Library, University of Wisconsin, Madison, Wisconsin.

* * * *

APPENDIX A

THE STATISTICAL ANALYSIS PROCEDURES

Statistical Analysis

The chief problem of prognosis to which research workers have given attention is the question of the predictive value of proposed measures. In treating any research relative to vocational success, prediction, and educational and vocational achievement, it is necessary to note certain matters pertaining to the meaning of test scores, the techniques of investigating the predictive value of measures, and the procedures for making predictions.

A. Meaning of Scores

Test scores are usually first expressed in terms of raw scores which in and of themselves are meaningless. A score of 145 on the Wechsler-Bellevue Intelligence Scale, for example, cannot be interpreted unless some additional information is given. It may be the lowest, the highest, or any score between two extremes for a particular group. Raw scores derive meaning through various forms of transmutation. They may be interpreted in terms of ranking whithin a group as the first or the fifth in a group of twenty.

More frequently the percentile rank is used, which indicates roughly the percentage of the group that falls below a given score or the position within a theoretical group of 100. Thus a percentile rank of 60 reveals that 60 percent of the group have lower ratings.

Scores derive meaning from coarse descriptive scales, for example, the Wechsler-Bellevue Intelligence scale has categories such as genius, superior, excellent, above normal, normal, below normal, and so on down the scale. Other forms of derived scores include age score, grade score, standard score, and the T-score.

B. Coefficient of Correlation

"The basic concept which has made prediction possible and which is the basis for evaluating the efficiency of prediction is that of correlation. By correlation is meant the relationship existing between capacities, traits, abilities, aptitudes, or various mental and physical functions. For example, in guidance it is very important to know as clearly as possible the relationship between the predictive items and the success criteria. This relationship is indicated best through the correlation coefficient r. The Pearson product-moment r is the best expression of this correlation. Its use is subject to certain limitations. For instance, this r should be used only with linear or fairly linear data and with data which may be considered continuous." (5)

In the analysis of the data in this study the Pearson product-moment r is used. The correlation coefficient changes in value from +1.00 to -1.00. The +1.00 indicates a perfect positive relationship and -1.00 indicates a perfect negative relationship. Practically all coefficients with educational and psychological material will be found to be significantly less than these two extremes.

By the fact that the correlation coefficient is used to show this prognostic relationship, it is not implied that this is the only method showing the relationship. It is, however, the one most commonly used and one which is the best expression mathematically to show the relationship.

"The general form of Pearson's product-moment coefficient correlation for deviations from the true mean is

$$r = \frac{x_1 x_2}{N_{1\ 2}}$$

where "r" is the coefficient correlation; x_1, x_2, is the summation of the product of the deviations of each measure from their true means; $_1$ and $_2$ are the standard deviations of the x_1 and x_2 distribution; and N is the number of cases used." (103)

C. Standard Error and Probable Error of the Coefficient of Correlation

The formula for the standard error of the coefficient of correlation is:

$$r_{12} = \frac{1 - r^2\ 12}{\sqrt{N}}$$

or when N is small

$$r_{12} = \frac{1 - r^2\ 12}{\sqrt{N - 1}}$$

This is the formula used when one wishes to test the null hypothesis, that is, to test the probability of getting a given coefficient of correlation by chance if the true correlation is 0. A correlation must be twice the size of its standard error in order to be significant at the 5% level. This means that there are only five (5) chances in one hundred of obtaining as large a correlation as that obtained by random fluctuations in sampling. This procedure is not valid for high values of r even though the sample is large. When the true correlation approaches ±1.00, the sampling distribution of the r's obtained even from large samples will be markedly skewed, and any interpretations of the standard error based on the normal probability integral table may be very seriously misleading.

The probable error of the coefficient of correlation will therefore be:

$$\text{P.E. } r\ 12 = .6745\ \frac{1 - r^2\ 12}{\sqrt{N}}$$

"If, for example, a correlation coefficient of .58 is obtained between two measures using 90 cases the P. E. will be equal to:

$$.6745\ \frac{1 - (.58)^2}{\sqrt{90}} = .05 \quad \text{and}$$

the coefficient correlation may be written .58 ± .05. This indicates that the chances are even that if another correlation coefficient is calculated upon 90 cases drawn from the same population as the first was drawn that the coefficient of correlation will not vary more than five points lower or higher than .58."(62)

D. Coefficient of Multiple Correlation

The multiple correlation coefficient is the correlation between observed values of one trait and values of that trait estimated by a regression equation from a multiple set of other traits. It may have any value between zero and +1, but it is never smaller than the largest of the zero order or partial r's involved in its computation.

The calculation as indicated here by the conventional formula is laborious,

$$E = 1 - 1 - r^2$$

when the number of dependent variables is greater than two or three, but a number of economical procedures have been developed. The above formula is used purely for predictive purposes.

APPENDIX B

THE RATING SCALE

Nature of the Rating Scale

The procedures and devices for the systematic recording of observations and judgments on a scale of units or values to be awarded to objects, persons, personality traits, attitudes, and the like may be conveniently classified into three categories: rating scales, rank-order method, and paired-comparison method. Of these, the rating scale is most frequently used and applied in educational investigations.

Research findings have been reported for the following problems:

(a) What traits are most amenable to rating?
(b) What number of scale steps or points can most observers discriminate in making the usual ratings?
(c) What types of rating scale are most valid and reliable for various educational purposes?
(d) What is the effect upon ratings of the influence of over-and-under estimation, central tendency, halo effects, and logical error?

The author has not attempted to review all the numerous studies that concern ratings, as this would require a volume in itself. We have selected a few which are illustrative of the method.

Adams, Furniss and De Bow (1) have done one of the most thorough and careful rating studies. Students were rated on 63 traits by ten or more associates. Traits showing a second-rating reliability of less than .70 were discarded from the study. For an Intelligence test the Army Alpha was used. The average correlation of grades with the ratings was much higher than the average correlation of the intelligence test with the ratings.

Pressey (83) found a correlation of .69 between "school attitude" (industry and application) and school marks. The same ratings correlated .55 with intelligence. Estimated "mental ability" correlated with grades .81. Sangren (92) had teachers rate their students on "scholarship qualities"(methods of work, application, industry and attitude, assimilation of new ideas), and correlated the scores with grades. The coefficient was .93. Inasmuch as the above factors are the prime determiners of most teacher's marks, the correlation should have been .99.

Fish (26), in a study of Wisconsin sorority women, found that the correlations of grades and rated personal characteristics was very low. The rating method used required that each rater rank the entire group, from first to last, thus eliminating the tendency to give too high ratings, etc. The students' honor points correlated as follows with the five traits rated: nervousness, -11; optimism, .28; industry in study, .38; physical activity, .28; leadership, .14.

These studies show that ratings can be used or misused, according to the amount of care exercised with the rating technique. We cannot emphasize too much the pitfalls into which the user may stumble. Many studies report correlations with ratings of personal traits which are terribly high; for instance, Webb (114) found a correlation between marks and "tendency to do kindnesses on principle" of .98. This means that the rater simply identified the two characteristics.

A trait is considered amenable to rating when competent judges (raters) tend to agree in their evaluations. Although verification is needed, findings from several studies indicate that some traits are more amenable to rating than others. Hollingsworth (48) found close agreement among raters upon such traits as efficiency, originality, perseverance, quickness, judgment, clearness, energy and will. There was fair agreement upon mental balance, leadership, intensity, reasonableness, independence, and physical health. There was poor agreement upon such traits as courage, unselfishness, integrity, cooperativeness, cheerfulness, and kindness. Shen (97) reported the best agreement upon scholarship, leadership, intelligence and the poorest upon punctuality, tact, and judgment. Miner (70) found high agreement for such traits as leadership, general ability, reliability, dependability, and sense of humor.

"The possibility of satisfactory agreement among judges appears to depend upon: (a) an adequate definition so that the judges are in agreement in regard to what is being rated, and (b) frequent occurrence so that the judges can observe sufficient samples of behavior. When the following traits are adequately defined and observed over a sufficient period of time, they are amenable to rating. They include: efficiency, originality, perseverance, quickness, emotional stability, interest, health, punctuality, leadership, scholarship, and intelligence. Such traits as courage, unselfishness, cheerfulness, kindness, judgment and tact have not been found amenable to reliable rating." (30)

No set rule can be laid down for the number of steps or units on a rating scale. Too few steps result in coarse ratings because the observer is not given opportunity to exercise fully his powers of discrimination, and too refined a scale makes it difficult for an observer to discriminate between one step and the next.

Conklin and Sutherland (16) concluded "that the maximum number of steps would be five for a single scale which extends from zero to a maximum and nine for a double scale which extends through zero with opposite qualities at the extremes of the scale."

Symonds (104) maintains that the degree of reliability one wishes in the final ratings should decide the matter. He concluded that for rating human traits seven is the optimal number. Less than seven leads to an appreciable loss in reliability.

Types of Rating Scales

"There are five major types of rating scales:

1. A number of phrases descriptive of varying degrees of a trait or characteristic arranged in order form a descriptive rating scale. In using this scale, an oberver selects the phrase that best describes the individual being measured.

2. If the descriptive phrases are printed at appropriate positions under a straight line, the instrument is called a graphic rating scale. (The scale used in this study is of this type.)

3. A product scale consists of a series of products, e. g., specimens of handwriting, arranged according to values determined by a group.

4. A man-to-man rating scale consists of descriptions of a number of persons, each of which is known to the user of the scale. This instrument has been widely used by the United States Army for rating officers.

5. A numerical rating scale, frequently called a score card, consists of a number of items (characteristics), each of which has been assigned a numerical value. '' (71)

Freyd (29) concluded that the graphic type of rating scale is the most satisfactory. He listed several advantages of this type of rating scale such as ''simplicity, comprehensibility, ease of administration, freedom from direct quantitative terms, and ratings as finely discriminated as the rater chooses.

Reliability and Validity of Ratings

Reliability of ratings may be increased by pooling the judgments of a number of persons, and validity increases with the objectivity of the trait or objected rated. The number of judgments that should be pooled (averaged) varies according to the degree of reliability sought and the nature of the trait rated. For many traits three or more independent ratings should be obtained. Rugg (90) recommends the use of the pooled ratings of not less than three independent judges; Symonds (104) recommends eight judges; and Bradshaw (9) from five to twenty-two judges. No one formula will apply to all situations. Furfey (33) has shown that the reliability of a rating may be increased by subdividing a trait into a number of subtraits and asking each judge to rate the subdivisions.

Cause of Errors

The types of errors which limit both the reliability and validity of ratings are:

1. The error of lenience, or severity, that is, rating all individuals too high or too low (frequently referred to as ''personal bias'' or ''personal equation'').

2. The error of central tendency, or hesitation to give ratings at the extremes of a scale.

3. Halo effect, or the general mental attitude toward or general impression of the person or object being rated.

4. The so-called logical error which comes from pre-suppositions in the minds of raters and lack of definiteness of the trait rated.

Personal bias in rating, such as a rater's tendency to overestimate or underestimate in judging a trait, introduces a systematic error. Conrad (17) pointed out that personal bias (personal equation) is not important in correlation studies, but in other situations the resulting errors may materially influence the findings. Guildford (40) proposed a formula to correct for this systematic error.

The halo effect was first mentioned by Wells (116), given its name by Thorndike, and described by Rugg (90), who remarked that we judge our fellows in terms of a general mental attitude toward them which leads us to attribute the general attitude or impression to particular qualities. The halo effect is likely to influence the ratings of a trait not easily observable or not clearly defined, such as those functioning in reactions with other people.

Ratings of the so-called ''traits of character'' or of personality traits of therapists are likely to involve a halo effect. Newcomb (73) pointed out a so-called logical error, attributed to logical presuppositions in the minds of the raters. This may be mitigated by calling for judgments of obviously observable actions rather than for abstract traits which may overlap.

Concluding Statements

The application of rating methods should be confined to traits amenable to rating. The rating of traits which are inferred from behavior should be based upon systematic observation over an extended period.

More precise and objective definitions of traits should be provided. Single ratings should not be used as measures of human qualities because sufficient reliability is usually obtained only when a composite of from five to ten independent observations are obtained. Practice in the use of a rating method will increase the reliability of the measures. So far as possible, bias should be eliminated from ratings.

For certain characteristics, such as personality traits, the rating scale has a definite place, but it is subject to errors and biases. As it is commonly used, it is heavily dependent on memory. Other difficulties arise in getting raters to agree upon the meanings of items as well as the degree of their presence in the situation or individual under observation. Rating scale techniques have been refined in the past through research. It is reasonable to expect that the rating scale will continue to be refined in the future.

APPENDIX C

DATA FORM

Tests and Their Uses

Tests are concerned with specific aspects of the individual, such as intelligence or mental ability, personal judgment, mechanical aptitude, interests, school knowledges, and the like. They should only be used when their administration and scoring can be done with competence. They must be used with care and regarded as important tools in the guidance process, but not substitutes for it. No single set of tests will solve all the problems of the guidance counselor.

Test may function in a number of ways as aids to counseling. They may be used to a limited extent both negatively and positively as instruments of prediction. If the counselor can determine what an individual has done on this or that test, he can then turn to tables of probabilities, and tell him that his chances of, say, graduating from college, or of learning the rudiments of electrical work, are practically nil. On the other hand, he can use test results positively and point out that, in the light of the counselee's demonstrable aptitude in certain areas of endeavor, he may choose from a whole range of job opportunities that fall in these fields, with a high probability that he will succeed in any of them provided his interest is great enough.

Tests have great value as diagnostic devices. A counselor may discover that an individual has definite aptitude for a certain vocation but apparently lacks special training for it. The question then arises: What are the inadequacies in this individual's program as it has developed up to this time? Diagnostic testing will greatly assist the counselor in answering this question, and thus enable the individual to maximize the training that he has had up to that time, and to concentrate upon the things he needs to know or do in order to develop any special competence needed to assure job success. Certain tests have a subsidiary purpose in this connection. The instructional test, for example, can be used to indicate how well an individual can learn by himself. This information is important for the counselor to know before he can make any recommendation to an individual being counseled about advancing his education.

Tests should be looked upon as supplementing each other and the counseling interview. Never take each test result and interpret it as if nothing else were known about the individual. Tests must be considered in relation to each other—which is why the "profile" is so important in the interpretation of test results. A counselor must get a great deal of practice with "profiles" before he can make a reasonable interpretation of them. Experience enables one to better develop the ability to perceive the patterns that tend to emerge.

APPENDIX N

CASE SUMMARY

1. Age. Miss Jane Huston was born August 19, 1936, at Mauston, Wisconsin, and now is twenty-one years of age.

2. Family status. Jane is the oldest of four girls, two of whom are attending a school of nursing and another doing office work. Jane has no brothers. Her father and mother are American born citizens who teach school in Milwaukee, Wisconsin.

3. Health status. At present her health is good. She had the usual childhood diseases, but there are

no evidences of complications or sequelae.

4. School status. Jane attended the elementary and secondary schools of Mauston. Three years ago she started at the University of Wisconsin as a freshman in Occupational Therapy course. Jane's academic average for three years is 1.2. In high school Jane ranked 24th out of fifty-eight students in her graduating class.

5. Reading habits. Jane reported to her advisor, Miss Caroline Thompson, that she reads and studies about two hours a day. Her scores on the Michigan Speed of Reading and the Wide Range Achievement tests are the 37th and 44th percentiles, respectively.

6. Extra-curricular activities. In high school Jane wrote for the monthly high school paper. She was active in dramatics. In college Jane engages in the social functions of the University Memorial Union.

7. Work history and odd jobs. Jane never had a job. She did baby-sitting on weekends for moderate remuneration for about three years.

8. Test results. Her percentile rank in the Henmon-Nelson test was the 40th.

	Score	Percentile
Intelligence		
Wechsler-Bellevue		
Verbal	91	41st
Performance	94	42nd
Total	93	42nd
California Mental Maturity		
Verbal	96	45th
Performance	99	48th
Total	98	48th
Achievement		
Wide Range Achievement Scale		
Reading	64	48th
Spelling	51	36th
Arithmetic	46	34th
Total	161	44th
Michigan Speed of Reading	42	37th
Interest		
Kuder Preference Record		
Mechanical		42nd
Computational		93rd
Scientific		17th
Persuasive		38th
Artistic		25th
Literary		51st
Musical		88th
Social Science (Service)		47th
Clerical		91st

Special Aptitudes

Purdue Pegboard	
Right Hand	56th
Left Hand	39th
Both Hands	48th
Total	46th
Rate of Manipulation	
Placing	45th
Turning	51st
Mechanical Comprehension Test	34th
Practical Judgment Scale	46th

Personality

Multiphasic Personality Inventory	
Cannot Say	50th
Lie Score	54th
F Score	50th
Hypochondriasis	56th
Depression	58th
Hypomania	53rd
Psychopathic-Deviate	56th
Masculinity-Femininity	58th
Paranoia	43rd
Psychasthenia	37tn
Schizophrenia	45th
Hysteria	37th
Washburne Social Adjustment Inventory	
Truthfulness	78th
Happiness	76th
Alienation	94th
Sympathy	90th
Purpose	87th
Impulse-Judgment	93rd
Control	89th

9. Diagnosis and evaluation. Jane is a student of average intelligence. Her scholastic attainment is below average. She ranks below the median intelligence quotient of the 234 students in Physical Medicine.

The test results clearly indicate that Miss Huston is stronger in non-language factors. Her speed of reading and comprehension is low. In both achievement instruments Jane ranked below average in reading and vocabulary. The low percentiles reveal that she does not have the ability to read scientific and medical literature efficiently and that her rate of reading is slow.

The profile record shows that Jane has little or no interest in the mechanical, scientific, artistic, and social service areas which is unlike the expected interest pattern of students interested in occupational therapy.

Both personality measures administered to Jane found her as average. Her responses indicate that Jane has a normal personality.

Her profile of the special aptitudes shows that Jane has none of the special aptitudes and skills expected of a good therapist. She had trouble with tip-of-the-finger movements, finger movements, and gross movements of the hands.

Recommendations

Jane was administered some positive and sincere guidance by the Department of Physical Medicine after all this data was perused and discussed with Jane by her advisor, Miss Thompson. It was suggested that she leave the Department because of low scholarship, lack of special aptitudes, lack of or a change of interest, poor reading skills, and because of her interest in the clerical and computational field.

Miss Thompson advised Jane to get into the Commerce school for a try-out or exploratory experience. A copy of this summary was sent to her advisor in her new field. Most recent reports have Jane enjoying her work. She seems to be better adjusted and is doing well.

TESTING THE ASSUMPTIONS UNDERLYING MULTIPLE REGRESSION

FRANCIS A. C. SEVIER
Glassboro State Teachers College
Glassboro, New Jersey

Introduction

THERE ARE, as all statisticians know, various assumptions underlying the application of multiple regression techniques. However, mention is seldom made by researchers who employ these techniques as to whether their assumptions have been tested. One must, therefore, take for granted, when reading such papers, that the researcher has cautiously analyzed each assumption.

One field in which multiple regression techniques have found wide application is in the prediction of academic success. In this field of application more sins have been committed in the name of multiple regression than perhaps anywhere else. This author, having become involved in a study of academic success, decided to see how well the data used in such educational problems really fit the assumptions.

The problem which was originally proposed to the author was that of predicting academic success in the Colleges for Men at Rutgers University. Freshman grade-point average (X_4) was to be predicted from the Verbal Aptitude Test score on the College Entrance Examination Board examination (X_1), the Mathematics Aptitude Test score (X_2), and rank in secondary-school graduating class (X_3). Sophomore cumulative grade-point average (X_5) was to be predicted from VAT score (X_1), MAT score (X_2), rank in secondary school (X_3), and freshman grade-point average (X_4). Junior cumulative grade-point average (X_6) was to be predicted from VAT score, MAT score, rank in secondary school, and sophomore cumulative grade-point average. Senior cumulative grade-point average (X_7) was to be predicted from VAT score, MAT score, rank in secondary school and junior cumulative grade-point average. Problems of this type are readily handled by multiple regression techniques.

The problem, when formulated mathematically, becomes one of writing the equation

$$Y_i = a_o + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

subject to a least-square criterion. (The b_n must be chosen in order to make the sum of the squares of the differences between the observed and estimated values of Y, a minimum.) In the case at hand this is a matter of writing Y_i ($i = 1, 2, 3, 4$), weighted cumulative grade average, for each of the six schools of the University, as a linear combination of the X_n ($n = 3, 4$): VAT score, MAT score, percentile rank in secondary-school graduating class, and weighted cumulative grade average through the previous undergraduate year. This notation implies that $n = 3$, $i = 1$ is the case of predicting freshman cumulative grade average, while $n = 4$, $i = 2, 3, 4$ are the cases of predicting sophomore, junior, or senior cumulative grade averages.

From the foregoing it is evident that the complete solution would require four regression equations for each of the six schools, or twenty-four regression equations in all. These twenty-four equations would permit the prediction of future academic status from VAT score, MAT score, percentile rank in secondary-school graduating class and current cumulative grade-point average. (All of these prediction possibilities, however, are predicated on the assumption that various statistical significance tests will be met.)

Before a complete regression analysis can be performed certain assumptions concerning the original data must be made and, where possible, tested, viz:

1. The X_n are fixed variates and may be looked upon as population parameters.
2. For a fixed set of X's, the Y's associated with this set are normally and independently distributed.
3. For any set of X's, the variance of Y shall be the same (this is the assumption of homoscedasticity).[1*]

Another assumption which is not listed in the above three, but which is perhaps the most important one in the regression analysis, must also be tested, viz., the linearity of regression of any X_n on Y or on any other X used in a particular regression equation. This assumption was tested by the familiar statistic η^2 (Eta square). Assumption (1) is in the nature of a postulate and must be made in order to satisfy the requirement of reliability of the X_n. The first part of assumption (2) was tested by using the X^2-test (chi square test) of normality. However, the normality of the marginal distributions does not insure normality in (2) unless regression is linear and the condition of homoscedasticity is fulfilled. Hence, assumption (3) was tested by using the procedure developed by Hartley for testing the homogeneity of variance.[2]

---

*All footnotes will be found at end of article.

One remark on notation must be clarified, the notation $(X_i, X_j; X_k)$ is employed to mean that the variables X_i and X_j are under consideration for all students who completed the kth year. Correspondingly, $(X_i; X_k)$ is employed to mean that the variable X_i is under consideration for all students who completed the kth year.

Linearity of Regression

Since 216 zero-order intercorrelations were involved in the complete study, it was decided to choose a representative sample of correlation tables to which tests of linearity of regression could be applied. Inferences concerning the assumption of linearity of regression were made from the results obtained from this sample. The rationale for the selection of the sample was as follows:

1. All permissible combinations of the seven variables involved in the study $(X_1, X_2, X_3, X_4, X_5, X_6, X_7)$ except (X_4, X_5), (X_5, X_6), and (X_6, X_7) were considered; there were fifteen such combinations. That the regressions of X_4 on X_5, X_5 on X_6, and X_6 on X_7 were linear could be seen from the correlation tables and the extremely high values obtained for the correlation coefficients: $r > .88$ in all cases.

2. By inspecting the correlation tables for particular combinations of variables, for all undergraduate levels and admission categories (six schools of the University), it was found that families of tables corresponding to the fifteen combinations in (1) could be formed on the basis of degree of linearity of regression. Thus the correlation tables were arranged in fifteen groups.

3. One member of each of these fifteen families was selected at random. However, no duplication of admission category and undergraduate level was permitted so that a wide sample could be achieved.

4. Nine combinations of undergraduate level and admission category would remain untested by the above method; hence, the correlation table showing the most apparent nonlinearity, for each of these nine combinations, was chosen to be tested for linearity of regression.

The statistic E^2_{yx} (sample estimate of η^2_{yx}) was computed for each of the twenty-four correlation tables.

The null hypothesis, that the true regression is linear, was tested by comparing the computed value of

$$F = \frac{E^2 - r^2}{1 - E^2} \quad \frac{N-6}{b-2}$$

for each of the twenty-four correlation tables with the entries in Snedecor's table for $n_1 = b - 2$ and $n_2 = N - b$ degrees of freedom (where N is the number of cases in an individual correlation table and b is the number of arrays in the x-direction). A significant value of F indicated a significant departure from linearity. Only one of the correlation tables displayed a significant departure from linearity at the 5% level.

Table I gives the summary data for the tests of linearity of regression. Since only one of the twenty-four correlation tables tested, $(X_1, X_3; X_4)$ in the Bachelor of Arts program, showed a significant departure from linearity at the level indicated, the assumption of linearity of regression of any X_n on Y or on any other X used in any regression equation in the study was considered tenable.

Normality

In order to test the normality assumption, it was not necessary to employ so elaborate a scheme for selecting distributions to be examined as that used in testing linearity of regression. The structure of the study reduces the admissible distributions for this test. For example, the marginal distributions of X_2 in all correlation tables for a particular undergraduate level and admission category are identical since the same grouping scheme was employed throughout the study. Hence, the 216 correlation tables reduce to 114 marginal distributions, nineteen for each admission category. That this reduction in admissible distributions can be made is apparent when it is recalled that four variables are involved in the freshman year, five in the sophomore year, five in the junior year, and five in the senior year for each admission category; and the grouping scheme is constant.

In order that the normality of each of the seven variables $(X_1, X_2, X_3, X_4, X_5, X_6, X_7)$ could be tested for each admission category, it was decided to test all variables in the freshman year, the variable X_5 in the sophomore year, and all variables in the senior year. Thus, the distributions of X_1, X_2, X_3 and X_4 were tested in the freshman year, the distribution of X_5 in the sophomore year, and the distributions of X_1, X_2, X_3, X_6, and X_7 in the senior year. In all, this amounted to 60 tests of normality.

The normality of a given marginal distribu-

TABLE I

TEST OF LINEARITY OF REGRESSION—STATISTICAL SUMMARY

Admission Category	Variables	E^2	E^2-r^2	$1-E^2$	n_1	n_2	F
Agriculture	$(X_2, X_4; X_4)$	.1543	.0483	.8457	17	231	0.7759
	$(X_3, X_5; X_5)$	.1076	.0314	.8924	15	193	0.4529
	$(X_1, X_6; X_6)$	.1455	.1014	.8545	10	146	1.7330
	$(X_1, X_3; X_7)$	.0936	.0779	.9064	8	133	1.4281
Bachelor of Arts	$(X_1, X_3; X_4)$	.0816	.0418	.9184	8	374	2.1271
	$(X_2, X_6; X_5)$	.0888	.0289	.9112	11	315	0.9078
	$(X_1, X_5; X_6)$	.2124	.0463	.7876	11	237	1.3203
	$(X_1, X_7; X_7)$	.2261	.0505	.7739	9	214	1.5527
Bachelor of Science	$(X_3, X_4; X_4)$	.2736	.0825	.7264	18	211	1.3316
	$(X_2, X_4; X_5)$	.1895	.1018	.8105	14	168	1.5072
	$(X_2, X_1; X_6)$	.1258	.0854	.8742	13	115	0.8643
	$(X_3, X_7; X_7)$	.1835	.0539	.8165	10	105	0.6930
Business Administration	$(X_2, X_3; X_4)$	.0656	.0503	.9344	8	250	1.6812
	$(X_1, X_5; X_5)$	.1297	.0708	.8703	18	180	0.8140
	$(X_2, X_6; X_6)$	.2036	.1475	.7964	15	118	1.4509
	$(X_3, X_6; X_7)$	.1844	.0006	.8156	10	105	0.0063
School of Education	$(X_1, X_3; X_4)$	.1143	.0578	.9422	7	109	0.9545
	$(X_1, X_4; X_5)$	.2004	.1072	.7996	14	88	0.8429
	$(X_3, X_6; X_6)$	.2514	.1394	.7486	10	72	1.3406
	$(X_2, X_7; X_7)$	.1368	.0685	.8632	8	63	0.6253
Engineering	$(X_2, X_1; X_4)$	.1945	.0335	.8055	15	192	0.5325
	$(X_1, X_3; X_5)$	.1990	.0397	.8010	15	163	0.5390
	$(X_3, X_5; X_6)$	.1508	.0820	.8429	10	137	1.3234
	$(X_1, X_6; X_7)$	.1048	.0562	.8952	10	124	0.7787

TABLE II

VALUE OF χ_0^2 AND CORRESPONDING P FOR TESTS OF "GOODNESS OF FIT" OF THEORETICAL NORMAL DISTRIBUTION TO OBSERVED DISTRIBUTION

	Agriculture			Bachelor of Arts			Bachelor of Science		
Variable	n	χ_0^2	P	n	χ_0^2	P	n	χ_0^2	P
$(X_2;X_4)$	8	9.94	$.30>P>.20$	10	10.93	$.40>P>.30$	10	20.49	$.05>P>.02$
$(X_1;X_4)$	7	10.54	$.20>P>.10$	12	5.78	$.95>P>.90$	11	7.10	$.80>P>.70$
$(X_3;X_4)$	5	11.88	$.05>P>.02$	5	37.04	$P<.001$	4	30.08	$P<.001$
$(X_4;X_4)$	12	14.55	$.30>P>.20$	13	18.35	$.20>P>.10$	12	21.20	$.05>P>.02$
$(X_5;X_5)$	10	17.17	$.10>P>.05$	9	26.28	$.01>P>.001$	11	8.08	$.80>P>.70$
$(X_2;X_7)$	8	8.80	$.40>P>.30$	9	10.52	$.40>P>.30$	9	7.32	$.60>P>.50$
$(X_1;X_7)$	7	5.96	$.60>P>.50$	10	8.02	$.60>P>.50$	8	11.65	$.20>P>.10$
$(X_3;X_7)$	5	13.21	$.05>P>.02$	4	25.43	$P<.001$	2	14.69	$P<.001$
$(X_6;X_7)$	6	39.26	$P<.001$	7	26.03	$P<.001$	8	11.29	$.20>P>.10$
$(X_7;X_7)$	5	17.77	$.01>P>.001$	7	26.86	$P<.001$	7	6.97	$.50>P>.40$

TABLE II
(Continued)

Business Administration			Education			Engineering		
n	χ_0^2	P	n	χ_0^2	P	n	χ_0^2	P
10	27.07	$.01>P>.001$	8	14.85	$.10>P>.05$	7	8.29	$.40>P>.30$
9	35.92	$P<.001$	8	3.83	$.90>P>.80$	10	12.00	$.30>P>.20$
5	23.60	$P<.001$	4	7.48	$.20>P>.10$	4	25.74	$P<.001$
14	10.87	$.70>P>.60$	9	19.93	$.02>P>.01$	12	14.35	$.30>P>.20$
9	26.82	$.01>P>.001$	7	6.26	$.60>P>.50$	7	39.86	$P<.001$
7	10.46	$.20>P>.10$	7	4.14	$.80>P>.70$	6	3.46	$.80>P>.70$
8	17.33	$.05>P>.02$	6	4.88	$.60>P>.50$	9	9.09	$.50>P>.40$
4	12.02	$.02>P>.01$	3	8.91	$.05>P>.02$	2	26.38	$P<.001$
6	20.17	$.01>P>.001$	4	6.25	$.20>P>.10$	7	9.04	$.30>P>.20$
5	18.39	$.01>P>.001$	4	2.83	$.60>P>.50$	6	13.33	$.05>P>.02$

TABLE III

TEST OF HOMOGENEITY OF VARIANCE—STATISTICAL SUMMARY

Variables	School	M_O	P
$(X_1, X_4; X_4)$	College of Agriculture	18.50	$P > .05$
$(X_2, X_4; X_4)$	Bachelor of Arts	9.69	$P > .05$
$(X_3, X_5; X_5)$	Bachelor of Arts	13.26	$P > .05$
$(X_3, X_4; X_4)$	Bachelor of Science	5.61	$P > .05$
$(X_3, X_4; X_4)$	Business Administration	33.03	$P < .01$
$(X_4, X_5; X_5)$	School of Education	13.14	$P > .05$
$(X_5, X_6; X_6)$	School of Education	18.51	$P > .05$
$(X_6, X_7; X_7)$	College of Engineering	3.64	$P > .05$

tion was tested by fitting a normal curve to the data, using the observed mean and standard deviation as estimates of the theoretical parameters, and then computing the corresponding χ_0^2 (chi square). The computed value of χ_0^2 was referred to a probability scale with $n = k - 3$ degrees of freedom; (where k is the number of classes in the marginal distribution) to test the null hypothesis, that the deviation of the sample from the theoretical normal distribution was due to sampling fluctuations. The following conditions had to be fulfilled in order to apply the χ^2 test:

1. No theoretical frequency should be very small.

2. The number of classes should not be very large.

It was therefore necessary to pool, as is standard procedure, the frequencies at the ends of the distributions so that no class had a theoretical frequency of less than five. (Cramēr[3] recommends pooling until the expected number in any class is at least ten. On the other hand, Cochran[4] has shown that, in some cases at least, numbers as small as "one" may be permitted without seriously affecting the validity of the test. It seems that pooling tends to increase the calculated probability, and hence to diminish the chance of rejecting the null hypothesis.)

A statistical summary of the χ^2 test for the 60 distributions tested is given in Table II. The value, P, in Table II is the probability of obtaining a value of χ^2 as large as, or larger than the observed χ_0^2. The hypothesis, that the underlying distribution is in fact normal, is to be rejected whenever $P < .05$.

With the range $P < .05$ assigned for rejection of the null hypothesis the following observations were made from Table II:

1. The distributions for the variables X_1 and X_2 refute the hypothesis of normality in four of twenty-four tests;

2. The distributions for the variable X_3 refute the hypothesis of normality in ten of twelve tests;

3. The distributions of the variables X_4, X_5, X_6, and X_7 refute the hypothesis of normality in eleven of twenty-four tests.

Thus twenty-five of the sixty tests refute the assumption of normality at the 5% level. Since the standard error of estimate is dependent upon the normality assumption for its interpretation, these results imply that the standard error of estimate will in any event have to be interpreted with caution.

Homogeneity of Variance

If the assumption of homoscedasticity could be justified, it would be possible to substitute one measure of error in place of the several, numerically different, array standard deviations. Thus the present section is devoted to testing the assumption of homoscedasticity which may be stated as follows:

Suppose there are b arrays in a correlation table, in the x-direction, with frequencies $N_1, N_2, \ldots, N_6$ and variances $s_1^2, s_2^2, \ldots, s_b^2$. The hypothesis is that $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_b^2 = \sigma^2$, i. e., the b samples have been drawn from normal populations having a common variance.

Three tests exist for testing this hypothesis: the L_1-criterion, used by Neyman and Pearson[5]; the Bartlett M-test[6]; and Hartley's[7] modification of the M-test. In the present study it was decided to use Hartley's test, since, as Johnson[8] points out, Bartlett's test is advantageous in comparison with L_1-test when the sample size is larger than sixty and since further, Hartley's test is more convenient as well as more accurate than the M-test. Tables for Hartley's test have been compiled by Thompson and Morrington[9].

Since one of the conditions of Hartley's test, namely the number of entries per array shall not be less than two, applied to only a few marginal distributions, it was decided to postpone the test, in the computational procedure, until the regression analysis had been completed. Of the forty-eight correlation tables remaining in the final regression analysis only four had marginal distributions fulfilling the aforementioned condition for the application of Hartley's test. Therefore, in order to achieve a more representative sample, four additional correlation tables were chosen at random from those remaining and the classes pooled to fulfill this condition of Hartley's test.

A statistical summary of the eight tests of the homogeneity of variance is given in Table III. The value of P in Table III is the probability of obtaining an M larger than the observed M_0. A sufficiently large M_0, one such that $F < .05$, resulted in a rejection of the null hypothesis of a common variance. From Table III it is apparent that in seven of the eight cases tested, the hypothesis of a common variance was tenable at the 5% level. Hence, the condition of homoscedasticity was considered fulfilled.

Summary

The analysis shows that for data such as VAT score, MAT score, and rank in secondary school graduating class one must be cautious in applying multiple regression techniques in the prediction of academic success.

The underlying assumption of linearity of regression seems to be tenable, as does the assumption of homogeniety of variance for the array distributions. The normality assumption, however, seems to be very much in doubt. Since the standard error of estimate is dependent upon the normality assumption for its interpretation, these results imply that the standard error of estimate will in any event have to be interpreted with caution.

Generally, these results indicate that more care must be exercised in employing multiple regression techniques than is currently the case. Since the analysis is based upon certain definite conditions or assumptions, it is imperative that these assumptions be analyzed.

FOOTNOTES

1. R. L. Anderson and T. A. Bancroft. Statistical Theory in Research (New York: McGraw-Hill Book Co., 1952), p. 168.

2. H. O. Hartley. "Testing the Homogeneity of a Set of Variances," Biometrika, XXXI (February 1940), pp. 249-255.

3. H. Cramer. Mathematical Methods of Statistics (Princeton, N. J.: Princeton University Press, 1946), p. 420.

4. W. G. Cochran. "The Chi-Square Distribution for the Binomial and Poisson Series, with Small Expectations," Annals of Eugenics, VII (1936), p. 215.

5. J. Neyman and E. S. Pearson. "On the Problem of k-Samples," Bulletin de l'Academie Polonaise des Sciences et des Lettres, Series A, (1931), pp. 460-481.

6. M. S. Bartlett. "Properties of Sufficiency and Statistical Tests," Proceedings of the Royal Society (London), Series A, CLX (1937), pp. 1-273.

7. H. O. Hartley. "Testing the Homogeneity of a Set of Variances," Biometrika, XXXI (1940), pp. 249-255.

8. P. O. Johnson. Statistical Methods in Research (New York: Prentice-Hall, 1949), pp. 82-86.

9. C. M. Thompson and M. Merrington. "Tables for Testing the Homogeneity of a Set of Estimated Variances," Biometrika, XXXIII, Part IV (1946), pp. 296-304.

THE INTERRELATIONSHIP OF VOCABULARY, COMPREHENSION AND RATE AMONG DISABLED READERS

JAMES C. REED and ROGER S. PEPPER
Wayne University

THERE HAVE been numerous scientific studies done to determine the nature of the reading act. Those investigations which have had as their aim the identification of parameters governing reading have frequently employed factor analysis. Studies which have been done on college populations, such as those of Pankaskie (9), Davis (2), Langsman (6) and Hall and Robinson (4), to mention only a few, have identified various facets of reading. The number of factors identified have differed according to the number and type of tests employed and the manner in which the factor analyses were carried out. However, there is general agreement that there is a vocabulary component, a factor involving comprehension of factual material, and one of speed of comprehension. By and large, these and similar studies have been done on a "normal reader" population. The subjects used in these studies have generally not had reading disabilities. While the factors of rate, vocabulary, and comprehension have been identified, to the writers' knowledge there have been no studies done on a college population to determine how these variables interact among persons having a reading disability. The reading clinician, however, does not see normal readers; his professional concern is with the deviates and he needs to know whether the factors which have been found to be present in reading operate in the same way for a deviate as they do for a normal reader.

The Problem

There are many ways of defining a reading disability. The reader is referred to any text on remedial reading (5, 8, 10). However, it has been defined by some as a discrepancy between listening comprehension and silent reading comprehension, and by some as a discrepancy between scores on an arithmetic test and scores on a reading test. Others have stated that a reading disability is present when there is a difference between some measure of intellectual capacity, such as the Stanford-Binet mental age, and some measure of reading level. All of the definitions subsume a disparity between manifest reading achievement and an estimate of potential reading ability. For the purpose of this study, a reading disability has been defined as extreme differences among the variables of rate, vocabulary and comprehension when these have been measured by the same reading test. It has been assumed that if a person makes a high score on any of these three factors, he has the potential ability to read well. One who scores low in all areas is considered to be a poor reader, rather than disabled. An individual who scores high in all areas is said to be a good reader. It is recognized that the foregoing definition of a reading disability, while methodologically sound, may have limited generality. Furthermore, a one-factor theory of reading is not a necessary concomitant of such a definition, but there is at least the implicit assumption that a general factor is present in reading. The problem of this investigation, then, is to determine the interrelationship of rate, vocabulary and comprehension among groups defined as having a reading disability. These reading disability cases are compared with groups who read well and groups who read poorly.

The Procedure

The Survey Section of the Diagnostic Reading Tests by Triggs et al., (11) was administered to approximately 2000 Wayne University students enrolled in freshman English. This test is divided into three parts: general reading, from which a rate score is obtained, vocabulary and comprehension. On the general reading section the student is required to read a story approximately 1800 words long and answer 20 questions about it. A total of 15 minutes is allowed for this section and the rate of reading is determined by the number of words read during the first three minutes. The vocabulary test has a ten-minute time limit and consists of 60 items. Each item is a definition followed by five words, one of which is correct, and the score is the total number right. The comprehension section has a fifteen-minute time limit and consists of four passages with five questions following each passage; but the comprehension score is the total number correct on this section plus the total number of questions answered correctly on the general reading section.

Local percentile norms were established for the freshman English students of Wayne. The

80th and 20th percentile ranks were used as cutting points. Subjects whose scores were at or above the 80th percentile rank were designated as high. Those whose scores were at or below the 20th percentile rank were called low. Three studies were made. For the first study, individuals were selected on the basis of their rate and comprehension scores. In the second study the individuals were chosen on the basis of their rate and vocabulary scores. In the third, the selection was made according to the vocabulary and comprehension scores. In each study the individuals were divided into four groups: (1) high-high, (2) low-low, (3) high-low, and (4) low-high. A person with a high-high classification is at or above the 80th percentile on two subtests. A low-low is a person at or below the 20th percentile on two subtests. A high-low person is one with a score at or above the 80th percentile on one subtest and at or below the 20th percentile on another subtest. A low-high person is one with scores just the opposite of a high-low's. Persons in these last two groups, high-low and low-high, are defined as having a reading disability. The other two groups are the good and the poor readers with whom the disability groups are being compared.

From each classification, 25 persons were selected by the use of a table of random numbers. N = 100 for each study. In general, each of the studies used different individuals. However, due to the random process of selection, some individuals appeared in more than one study.

Analysis of the Results

In each of the three studies, a two-way analysis of variance was used to determine the interrelationship of rate, vocabulary and comprehension. The effect of either or both of any pair of variables upon the variance of the third was studied. Determining interrelationships is also a problem for multiple correlation and multiple regression; however, these techniques would not be suitable in the present study. Due to the way in which the groups were classified, the marginal distributions would not be normal, hence, difficulty would have been encountered in making tests of significance had multiple correlation been used. That technique would be appropriate only if a random sample had been drawn from the entire population of N = 2000.

Since each of the categories in the two-way classification is fixed, Model I is appropriate, and the proper error term is the within cells variance. That this is so can also be seen by examining the expected values of the Mean Squares. (For a discussion of Models in the Analysis of Variance, see references 1, 3, 7, 12.)

The summaries of the two-way analyses are presented in Tables I, II, and III. The results will be considered separately for each analysis.

With vocabulary the dependent variable $\underline{F}$ was found to be significant at the .5% level for rate and significant beyond the .1% level for comprehension. No significant interaction was found.

In the second study, with comprehension the dependent variable $\underline{F}$ was found to be significant between the 5% and the 2% levels for rate, and significant beyond the .1% level for vocabulary. No significant interaction variance was found.

In the third study, where rate was the dependent variable, $\underline{F}$ was found to be significant beyond the .5% level for both vocabulary and comprehension, and, again, the interaction was not significant.

Discussion and Conclusions

The foregoing results need to be interpreted cautiously. The various scores on the Diagnostic Survey Test are correlated postively. As such the distributions of the scores beyond the points of truncation may not be normal and heterogeneity of variance may be present even though it was not detected by Bartlett's test. An asymmetrical distribution might be more likely to occur among the groups classified as normal, that is the low-low and the high-high categories, as these are the tails of a normal distribution. The number of cases falling within the disabled categories, low-high and high-low, was much fewer than in the low-low and high-high classifications. In fact, a total of only 26 out of the entire 2000 cases was found for the high comprehension-low vocabulary group. Due to the positive correlation of the variables, coupled with the smaller population numbers of certain categories and the way in which the groups were selected, it is possible that the assumptions underlying an analysis of variance have not been met. However, a visual inspection of the data did not reveal marked departure from normality.

Under the restrictions of the previous remarks, several implications can be drawn from the results. Of interest is the interpretation of the insignificant interaction. In general, non-significant differences mean only that a specified hypothesis cannot be rejected. It may be that no differences actually exist, or it may be that, due to the power of the test, a specified level of significance was not reached. In the situation at hand, the insignificant interaction does not mean that the two variables have zero correlation. The fact that in each of the two-way classification studies the main effects were all significant, with one possible exception, shows that the variables are related. Consider the exact nature of the model: $\xi_{ij} = \xi_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \overline{e}_{ij}$. ξ_{ij} is

TABLE I

MEANS, VARIANCES AND SUMMARY OF THE ANALYSIS OF VARIANCE OF VOCABULARY SCORES

Group n = 25	H_cH_r*	H_cL_r	L_cH_r	L_cL_r
Mean	49.12	42.80	35.24	32.24
Variance	54.54	66.00	97.77	45.44

Source of Variation	df	Mean Square	P	Expected Value of Mean Square
Rate	1	542.89	.005	$\sigma^2 + 50\ \Sigma R^2$
Comprehension	1	3733.21	$<.001$	$\sigma^2 + 50\ \Sigma C^2$
Interaction	1	68.89		$\sigma^2 + 25\ \Sigma(RC)^2$
Within Cells	96	65.94		σ^2

*Legend: $H \geq 80$ percentile c = comprehension

$L \leq 20$ percentile r = rate

TABLE II

MEANS, VARIANCES AND SUMMARY OF THE ANALYSIS OF VARIANCE OF COMPREHENSION SCORES

Group n = 25	H_vH_r*	H_vL_r	L_vH_r	L_vL_r
Mean	32.36	31.92	28.12	24.36
Variance	12.82	11.83	33.84	24.99
Source of Variation	df	Mean Square	P	Expected Value of Mean Square
Rate	1	110.25	$<.05$	$\sigma^2 + 50\ \Sigma R^2$
Vocabulary	1	870.25	$<.001$	$\sigma^2 + 50\ \Sigma V^2$
Interaction	1	68.89		$\sigma^2 + 25\ \Sigma(RV)^2$
Within Cells	96	20.90		σ^2

*Legend: $H \geq$ 80 percentile v = vocabulary

$L \leq$ 20 percentile r = rate

TABLE III

MEANS, VARIANCES AND SUMMARY OF THE ANALYSIS OF VARIANCE OF RATE SCORES

Group n = 25	H_cH_v*	H_cL_v	L_cH_v	L_cL_v
Mean	76.24	65.88	65.76	58.88
Variance	172.94	141.19	175.19	211.53

Source of Variation	df	Mean Square	P	Expected Value of Mean Square
Comprehension	1	1857.61	$<.005$	$\sigma^2 + 50\ \Sigma C^2$
Vocabulary	1	1909.69	$<.005$	$\sigma^2 + 50\ \Sigma V^2$
Interaction	1	75.69		$\sigma^2 + 25\ \Sigma(CV)^2$
Within Cells	96	175.21		σ^2

*Legend: $H \geq 80$ percentile　c = comprehension

$L \leq 20$ percentile　v = vocabulary

the true mean of the ijth cell. $\xi_{..}$ is the grand mean of the population. α is the deviation of the grand mean from the row mean ($\xi_{i.} - \xi_{..}$). β is the deviation of the grand mean from the column mean ($\xi_{.j} - \xi_{..}$). $\alpha\beta$ is the interaction ($\xi_{ij} - \xi_{.j} + \xi_{..}$) and $\overline{e}_{ij}$ is the error component. Assume $(\alpha\beta)_{ij}$ (the interaction) is 0. The cell mean can be expressed as the sum of two parts, one due to the "row effect" and the other due to the "column effect", together with a constant. If the interaction were significant, it would mean that some combination of the two factors would lead to an unfavorable result, others to a favorable result, as compared with the results calculated from the row and column means. In general, under Model I, when the interaction is significant, there is not much interest in testing the main effects, although there is no logical objection to doing so. When the interaction is 0, the effects of the two factors can be considered to be independent.

Apply the foregoing discussion to the relationship of the variables being considered. It can be said, for example, that while comprehension and vocabulary are correlated, their effects on rate are additive since the interaction is not significant. At least, that hypothesis cannot be rejected. In other words, the mean rate score for the ijth cell can be expressed in terms of a comprehension effect plus a vocabulary effect and a constant. To be more concrete, under the assumption of additivity, persons who are high in both vocabulary and comprehension would be expected to have a higher rate score than those high in one and low in the other.

If one is willing to assume a linear model with additivity, then the definition of a reading disability as used in this study appears to have some merit. For the purpose of screening individuals for further testing or possible short-term training, the classifications as suggested may have use. An individual high in one reading skill but low in another will not only require less time than one who is low in all, but he would be a safer bet for success. The latter, the "low-low" individual, may be a poor reader, but doing the best he can with generally limited abilities. Of course, with reading disability defined otherwise, he may be a severely retarded reader when his reading achievement is compared with his intellectual capacity, in which case he would probably require individual rather than group training.

Under the same assumption of additivity, it can be said that improving skill in one area of reading may contribute to proficiency in another. There may be reason to doubt the oft-repeated statement to the effect that all reading skills must be taught. Logically, however, whether or not improvement is obtained in skills other than the one being stressed will depend on the particular skill involved. Developing vocabulary may result in increased comprehension. However, rate training will probably not contribute to vocabulary, although it may result in a higher vocabulary score being obtained on a timed test.

To answer the original question posed in this investigation, the variables herein studied appear to operate among reading deviates in the same way as they do in a normal population. Furthermore, when two of the three variables are used as a basis for classification, subjects who are high on both will be high on the third. Subjects high in one but low in the second will tend to be higher on the third than those who are low in both. While rate, vocabulary and comprehension are related, the separate effects of any two of them on the third appear to be additive.

In addition to the interrelationships as discussed in the foregoing paragraphs, the following statements can be made about the variables themselves:

Vocabulary— The score on the vocabulary test is affected by rate. It is a timed test and to complete it within the specified limits, one must proceed at the rate of answering six questions a minute. However, judging only from the magnitude of the differences among the means, it appears that the time element affects those with good comprehension more than those with poor comprehension. The data suggest that for good comprehenders with a slow rate, the vocabulary score on this test may be an under-estimate of their actual word power. On the other hand, for those who are low in comprehension the vocabulary score may be a valid estimate of their word knowledge regardless of rate.

Comprehension—Of the variables studied, this one appears to be related only to vocabulary. True, F for rate was significant beyond the 5% level, but it is not believed to be important. Certainly, there is no important difference between the means of the high vocabulary-high rate group and the high vocabulary-low rate group. It may be that there is some small advantage for those who read rapidly in that they can answer the questions and have time to check their work, but the time element on the comprehension section does not seem to matter greatly. As would be expected, the comprehension score is closely related to the vocabulary score, and this relationship needs no explanation.

Rate—Perhaps the most interesting finding of this study is the relationship of rate to the variables of comprehension and vocabulary. Since rate is so much a function of vocabulary and comprehension, the question is raised as to whether it has a variance of its own. Perhaps the most important variables governing speed of reading, instead of lying in the perceptual domain, can be accounted for by such factors as rate of concept formation, organizational ability and general word knowledge. One may well wonder as to the justification of such training devices as reading films and tachistoscopes. The findings also suggest that it does not appear profitable to use rate as a measure of potential reading ability.

REFERENCES

1. Brownlee, K. A. Model III in the Analysis of Variance, unpublished manuscript, University of Chicago, 1955.
2. Davis, Frederick B. "Fundamental Factors of Comprehension in Reading," Psychometrika, IX (September 1944), pp. 185-97.
3. Eisenhart, Churchill. "The Assumptions Underlying Analysis of Variance," Biometrics, III (1947), pp. 1-21.
4. Hall, William, and Robinson, Francis. "An Analytical Approach to the Study of Reading Skills," Journal of Educational Psychology, XXXVI (October 1945), pp. 429-42.
5. Harris, Albert J. How to Increase Reading Ability, 3rd ed., (New York: Longmans, Green and Co., 1956).
6. Langsman, Rosalind S. "A Factorial Analysis of Reading Ability," Journal of Experimental Education, X (September 1941), pp. 57-63.
7. McNemar, Quinn. Psychological Statistics, 2nd ed., (New York: John Wiley and Sons, 1955), Ch., 16, pp. 281-342.
8. Monroe, Marian. Children Who Cannot Read (Chicago: University of Chicago Press, 1932).
9. Pankaskie, Margaret. "Some Factors in Reading Achievement at the College Level," Psychological Bulletin, XXXVII (October 1940), p. 573.
10. Robinson, Helen M. Why Pupils Fail in Reading (Chicago: University of Chicago Press, 1946).
11. Triggs, Frances, et al. Diagnostic Reading Tests, Committee on Diagnostic Reading Tests, Kingscote Apartment 3g, 419 West 119th St., New York 27, N. Y.
12. Wilk, M. B. and Kempthorne, O. "Fixed, Mixed and Random Models," Journal of American Statistical Association, L (December 1955), pp. 1144-67.

A QUICK AND DIRTY METHOD FOR ESTIMATING MULTIPLE CORRELATIONS

WILLIAM J. MOONAN*
USN Personnel Research Field Activity
San Diego, California

1. Introduction

IT IS WELL known that the lower bound of the estimate of the multiple correlation is the largest zero-order product moment correlation coefficient which exists between the criterion variable and the independent variables. Of course, the upper bound is unity. Frequently we desire to know a more accurate estimate than is provided by these two limits, particularly if they are quite divergent and if a satisfactory estimate can be easily obtained. This knowledge is valuable because we can use such an estimate to decide whether or not it is worthwhile to proceed with the evaluation of the multiple regression equation or the maximum likelihood estimate of the multiple correlation.

The purpose of this note is to report a procedure which enables a quick estimation of the multiple correlation for any number of independent variables. The estimate given here is very simply found if the zero-order correlations among the independent variables and with the dependent variable are available. The state of knowledge regarding the precision properties of the suggested estimate is, at present, purely empirical. Consequently, the method cannot be offered with as much statistical detail as desirable. However, the procedure has been tried with several examples and, in most cases, has provided estimates which are surprisingly close to the maximum likelihood estimates even under conditions which would indicate that they should be quite different. Of course, optimally, the real value of the estimate provided here should be assessed according to how well it estimates the population multiple correlation.

The author realizes that the act of showing that the method "works" with several sets of data does not prove that it is "good" in general. However, it does seem to add some degree of credibility to the procedure and this type of information is, at present, the best available. It should be emphasized that this empirical evidence, although encouraging, is not too substantial and anyone who uses the method is advised to regard his results accordingly. In other words, the estimate is "dirty", but the degree of this dirtiness is unknown.

2. Notation and Preliminaries

Capital letters, except for the symbols R and $\overline{R}$, will stand for matrices or vectors. A and $\overline{A}$ will denote non-singular, symmetrical, square intercorrelation matrices. The elements of A are r(i, j); i, j=1, . . . , k where r(i, i) equals 1 and r(i, j) for i$\neq$ j equals the product-moment correlation between the i and jth independent variables.

The elements of the main diagonal of $\overline{A}$ are 1 and all other elements are equal to the constant,

$$(1) \qquad \bar{r} = 2 \sum_{i=1}^{k-1} \sum_{j>i}^{k} r(i, j)/k\overline{k-1}.$$

That is, the non-diagonal elements of $\overline{A}$ are the average of the non-diagonal intercorrelations of A. We may term $\overline{A}$ a centro-iso-symmetric matrix since its non-diagonal elements are all equal to $\bar{r}$ and its diagonal elements equal unity. Thus, if we think of the matrix as describing a physical ellipsoid, it would have the same symmetrical properties in all directions with respect to the primary principal axis of the ellipsoid. The inverses of A and $\overline{A}$ are indicated by A^{-1} and $\overline{A}^{-1}$.

C will indicate a vector consisting of one row and k columns. Its elements will be the intercorrelations of the independent variables, x(i) and the criterion variable, y. These elements are denoted by r(i;y); i = 1, . . . , k. The transpose of C is indicated by C'. In this paper, primes will always indicate transposed vectors.

The maximum likelihood estimate of the multiple correlation, i. e., R, is given by the square root of the quadratic form $CA^{-1}C'$, that is

$$(2) \qquad R = (CA^{-1}C')^{\frac{1}{2}} .$$

The inverse of any centro-iso-symmetric matrix and, in particular $\overline{A}$, has been given by Roy and Sarhan (9). The main diagonal elements of $\overline{A}^{-1}$ are all equal and may be denoted by $g = \theta e$, and all non-diagonal elements are also equal and are denoted by $h = -\theta\bar{r}$, where $\bar{r}$ is defined by (1),

---

*The opinions expressed are solely those of the author and are in no way official; nor are they intended to be construed as representing those of the U. S. Naval Personnel Research Field Activity or the Bureau of Naval Personnel.

$$\theta = 1/[(1-\bar{r})(1+\overline{k-1}\;\bar{r})], \quad \text{and} \tag{3}$$

$$e = 1 + \overline{k-2}\;\bar{r}.$$

3. The Method and an Example

The procedure suggested here amounts to calculating $\bar{r}$ from A, finding the inverse of $\bar{A}$, and calculating

$$\bar{R} = (C\bar{A}^{-1}C')^{\frac{1}{2}}. \tag{4}$$

To evaluate the procedure, we will compare $\bar{R}$ to R as calculated from (2). To illustrate the calculations, consider the following information which has been taken from unpublished data. Let,

$$A = \begin{bmatrix} 1.000 & .183 & .323 & .246 \\ .183 & 1.000 & .651 & .537 \\ .323 & .651 & 1.000 & .623 \\ .246 & .537 & .623 & 1.000 \end{bmatrix}$$

and,

$$C = [.281 \quad .484 \quad .303 \quad .393]$$

In this $k = 4$ case, and because $\sum_{i=1}^{3}\sum_{j>i}^{4} r(i,j) = .183 + .323 + .246 + .651 + .537 + .623 = 2.563$, $\bar{r} = 2(2.563)/12 = .427$, $\theta = 1/(.573)(1+3(.427)) = 1/1.3070 = .765$, and $e = 1 + 2(.427) = 1.854$. The value of $g = \theta e$ is 1.418 and $h = -\theta\bar{r} = -.327$. It is recommended that the arithmetic be checked at this point by determining if $g + \overline{k-1}\,h\bar{r} = 1$. For this example, $1.418 + 3(-.327)(.427) = .999$, which is satisfactory when rounding errors are considered.

Therefore,

$$\bar{A}^{-1} = \begin{bmatrix} 1.418 & -.327 & -.327 & -.327 \\ -.327 & 1.418 & -.327 & -.327 \\ -.327 & -.327 & 1.418 & -.327 \\ -.327 & -.327 & -.327 & 1.418 \end{bmatrix}$$

If the matrix multiplication indicated by the triple matrix product $C\bar{A}^{-1}C'$, (4) is performed, we find $\bar{R}^2 = .278197$ so that $\bar{R} = .527$. The maximum likelihood estimate given by (2) is $R = .540$. So we see that in this case, the agreement between $\bar{R}$ and R is quite good.

The effort involved in multiplying $C\bar{A}^{-1}C'$ can be reduced considerably due to the special character of $\bar{A}^{-1}$. Note that in the process of multiplying the i^{th} column of $\bar{A}^{-1}$ by C, all elements of C except the element which occurs in the i^{th} column, are multiplied by h. The element in the i^{th} column is multiplied by g. Therefore, if we sum up the elements of C, i.e.,

$$\sum_{i=1}^{k} r(i;y) = \phi, \quad \text{say},$$

and subtract out the i^{th} element, we would have all the necessary values which need to be multiplied by h. These can be listed in the vector D, where

$$D = [\sum_{i=1}^{k} r(i;y) - r(i;y)] = [\phi - r(i;y)];\ i = 1, \ldots, k. \tag{5}$$

Then,

$$\bar{R}^2 = C\bar{A}^{-1}C' = [gC + hD]C' = \bar{B}C' \tag{6}$$

where,

$$\bar{B} = [gC + hD]. \tag{7}$$

For the example,

$$gC = 1.418 \quad [.281 \quad .484 \quad .303 \quad .393]$$
$$= [.398 \quad .686 \quad .430 \quad .557].$$

Since $\phi = \sum_{i=1}^{4} r(i;y) = 1.461$, the elements of D are $1.461 - .281 = 1.180$, $1.461 - .484 = .977$, $1.461 - .303 = 1.158$, and $1.461 - .393 = 1.068$. We find

$$hD = -.327 \quad [1.180 \quad .977 \quad 1.158 \quad 1.068]$$
$$= [-.386 \quad -.319 \quad -.379 \quad -.349].$$

Using equation (7)

$$\bar{B} = [.012 \quad .367 \quad .051 \quad .208]$$

and $\bar{R}^2 = \bar{B}C' = .278197$, giving $\bar{R} = .527$. In addition to the computational check, $g + \overline{k-1}\,h\bar{r} = 1$, it can be seen that the sum of the elements of the vectors gC, D, hD and B are respectively equal to $g\phi$, $\overline{k-1}\,\phi$, $\overline{k-1}\,h\phi$ and $\phi(g+\overline{k-1}\,h)$. These checks should be applied at the appropriate time in the course of the computations.

$\bar{B}$ was labeled as such because it approximates the standardized maximum likelihood regression coefficients which would be determined from the equation $BA = C$ by the equation $B = CA^{-1}$. The value of B, for this example is

$$B = [.207 \quad .457 \quad -.199 \quad .220].$$

This vector does not compare well with $\bar{B}$. Experience has shown that unless the $r(i, j)$ are rel-

TABLE I

THE VALUES OF R AND $\overline{R}$ FOR DATA FROM SEVERAL PUBLISHED SOURCES

Source of Data	Value of			
	k	R	$\overline{R}$	$\|R-\overline{R}\|$
Kelly (5)	3	.51	.52	.01
Mahalanobis, et al. (6)	3	.31	.32	.01
Snedecor (10)	3	.74	.74	.00
Tate (11)	3	.73	.72	.01
Walker and Lev (13)	3	.66	.68	.02
Griffin (2)	4	.76	.78	.02
Guilford (3)	4	.70	.71	.01
McNemar (7)	4	.74	.72	.02
Peters and Van Voorhees (8)	4	.73	.79	.06
Johnson (4)	5	.70	.90	.20
Waits (12)	5	.50	.50	.00
Wert, et al. (14)	5	.44	.38	.06
Greenberger and Ward (1)	20	.40	.43	.03

atively homogeneous, and therefore $\overline{A}$ is a good representation of A, the B and $\overline{B}$ vectors are not particularly equivalent. The values of R and $\overline{R}$ have appeared to be much more compatible. This is apparently due in part to the "averaging effect" produced by "errors" in the elements of B counterbalancing each other in the evaluation of $\overline{R}^2 = \overline{B}C'$. Even though $\overline{B}$ may not agree precisely with B, it may prove a useful starting point for methods such as the one given by Greenberger and Ward (1) for finding B by iteration. In passing, it is interesting to note that when $k = 2$, $\overline{R}$ will exactly equal R and $\overline{B}$ will exactly equal B, since $\overline{r}$ is equal to $r(1, 2)$. Also, for an experienced operator, who knows the procedure, it takes about two minutes per independent variable to compute $\overline{R}$ if A and C are given and an electric desk calculator is available.

4. Other Evidence

The procedure has been applied to data taken from several textbooks and published articles which deal with statistical methods or research data. These examples have been selected because the data were expressed in terms of product-moment correlation coefficients. This method of presentation was most useful for the application of the method suggested here. The results of the applications of the method appear in Table I. This table presents the source of the data, the number of independent variables, the values of R and $\overline{R}$ and their absolute difference.

Only for the example given by Johnson (4) was $\overline{R}$ in serious disagreement with R. Scrutiny of the correlation data given by Johnson will show that the independent variables cannot be represented very accurately by a five-dimensional centro-iso-symmetric matrix with an $\overline{r} = .293$ and so we should not expect very good agreement between R and $\overline{R}$. This particular example was included to caution the reader that the method should not be applied indiscriminantly.

The problem of determining whether or not the method can profitably be applied to a set of data depends, of course, upon the character of A and C and the value of k. The use of the mean deviation of the $r(i, j)$ suggests itself as a possible way of obtaining a quickly determined estimate of whether A can be adequately represented by $\overline{A}$. The more homogeneous are correlations in A, the better A can be represented by $\overline{A}$. Other factors neglected in this evaluation are the "pattern" of the $r(i, j)$ and the $r(i; y)$. These are as relevant as the magnitudes of the $r(i, j)$ themselves and probably account for the fact that the $|R-\overline{R}|$ have not appeared, in the examples on which the method has been tried so far, to be very related to the mean deviation of the $r(i, j)$. It is with considerable regret that guidance as to whether or not the method is useful for any particular problem must depend upon the user's judgment as to whether or not A is adequately represented by $\overline{A}$. The numerical examples show that considerable latitude in the decision is permissible, but caution must be exercised in doubtful cases.

From the results of examples worked from the published and other unpublished data, it is the author's judgment that the method holds reasonable promise for quickly providing a good estimate of multiple correlations providing that the $r(i, j)$ are not extremely heterogeneous.

5. Summary

A procedure, which is arithmetically simple, has been given for the purpose of providing a quickly obtained estimate of multiple correlations. It has been empirically shown that the estimate provided by the procedure is "practically" precise, in respect to the maximum likelihood estimate, in most cases. The user is cautioned that the method does not take into consideration the "pattern" of the intercorrelations among the independent variables, but merely utilizes their average value and their number. Other things being equal, it is expected therefore that the method becomes less exact as the heterogeneity of the intercorrelations among the independent variables increases. The method was applied to several sets of published data and the results of these applications are presented. They generally indicate that the method is practical in the sense of providing estimates which are nearly equal to the maximum likelihood estimates of multiple correlations. Statistical evidence relating to the mean and precision properties of the suggested estimate are, at present, unfortunately unknown. The user is cautioned accordingly.

An arithmetic procedure was shown whereby the evaluation of a triple matrix product used to provide the multiple correlation estimate, $\overline{R}$, could be reduced to the simple problem of multiplying two vectors together. To this end, the quantities involved and defined in the procedure should be listed and evaluated in the following order: $A, k, \overline{r}, \theta, e, g, h, C, gC, \phi, D, hD, \overline{B}, \overline{B}C'$, and finally $(\overline{B}C')^{\frac{1}{2}}$ which is equal to $\overline{R}$.

REFERENCES

1. Greenberger, M. H. and Ward, J. H. "An Iterative Method for the Solution of Regression Weights and Similar Problems," IBM Technical Newsletter, No. 6 (New York: Applied Science Division, IBM Corp., 1953), p. 66.

2. Griffin, H. D. "Simplified Schemas for Multiple Linear Correlation," Journal of Ex-

perimental Education, I (1933), p. 239.

3. Guilford, J. P. *Fundamental Statistics in Psychology and Education* (New York: McGraw-Hill, Inc., 1950), p. 427.
4. Johnson, P. O. *Statistical Methods in Research* (New York: Prentice-Hall, Inc., 1949), p. 332.
5. Kelley, T. L. *Fundamentals of Statistics* (Cambridge, Mass.: Harvard University Press, 1947), p. 479.
6. Mahalanobis, P. C. and Khagendra, N. C. "Studies in Educational Tests No. 3 Analysis of Marks in the School Leaving Certificate Examination in the United Provinces, India, 1919," *Sankhya*, I (1933), p. 254.
7. McNemar, Q. *Psychological Statistics* (New York: John Wiley and Sons, 1949), p. 158.
8. Peters, C. C. and Van Voorhees, W. R. *Statistical Procedures and Their Mathematical Bases* (New York: McGraw-Hill, Inc., 1940), p. 230.
9. Roy, S. N. and Sarhan, A. E. *On Inverting a Class of Patterned Matrices, Part I* (Chapel Hill: Institute of Statistics, University of North Carolina, 1955), p. 20.
10. Snedecor, G. W. *Statistical Methods* (Ames Iowa: Collegiate Press, 1946), p. 298.
11. Tate, M. W. *Statistics in Education* (New York: Macmillan Co., 1955), p. 307.
12. Waits, J. V. "The Differential Predictive Value of the Psychological Examination of the American Council on Education," *Journal of Experimental Education*, I (1955), p. 267.
13. Walker, H. M. and Lev, J. *Statistical Inference* (New York: Henry and Co., 1953), p. 327.
14. Wert, J. E. and others. *Statistical Methods in Educational and Psychological Research* (New York: Appleton-Century-Crofts, 1954), p. 250.